TRANSACTIONS

OF THE

AMERICAN MATHEMATICAL SOCIETY

EDITED BY

ROBERT D. CARMICHAEL

FRANCIS R. SHARPE

JACOB D. TAMARKIN

VOLUME 35
1933

PUBLISHED BY THE SOCIETY
MENASHA, WIS., AND NEW YORK
1933

General
Bound January, 1934

COMPOSED, PRINTED AND BOUND BY
The Collegiate Press
GEORGE BANTA PUBLISHING COMPANY
MENASHA, WISCONSIN

TABLE OF CONTENTS

VOLUME 35, 1933

PAGE

ADAMS, C. R., and CLARKSON, J. A., of Providence, R. I. On definitions of bounded variation for functions of two variables 824
AGNEW, R. P., of Providence, R. I. On Riesz and Cesàro methods of summability 532
ALBERT, A. A., of Chicago, Ill. Non-cyclic algebras of degree and exponent four 112
——— Cyclic fields of degree eight 949
BECKENBACH, E. F., and RADÓ, T., of Columbus, Ohio. Subharmonic functions and minimal surfaces 648
——— Subharmonic functions and surfaces of negative curvature . . 662
BELL, E. T., of Pasadena, Calif. The Latin square, or cyclic, functions . 734
——— Polynomial diophantine systems 903
BLISS, G. A., and HESTENES, M. R., of Chicago, Ill. Sufficient conditions for a problem of Mayer in the calculus of variations 305
BRAHANA, H. R., of Urbana, Ill. Groups $\{S, T\}$ whose commutator subgroups are abelian 386
BULLOCK, R. C., of Jackson, Tenn. Non-conjugate osculating quadrics of a curve on a surface 518
CAIRNS, S. S., of Bethlehem, Pa. An axiomatic basis for plane geometry . 234
CARLITZ, L., of Durham, N. C. On abelian fields 122
——— On the representation of a polynomial in a Galois field as the sum of an even number of squares 397
CARMICHAEL, R. D., of Urbana, Ill. Systems of linear difference equations and expansions in series of exponential functions 1
CLARKSON, J. A., and ADAMS, C. R., of Providence, R. I. On definitions of bounded variation for functions of two variables 824
CURRIER, A. E., of Cambridge, Mass. Proof of the fundamental theorems on second-order cross partial derivatives 245
DIAMOND, A. H., of Berkeley, Calif. The complete existential theory of the Whitehead-Huntington set of postulates for the algebra of logic . . 940
DOOB, J. L., of New York, N. Y. The boundary values of analytic functions. II. 418
GERGEN, J. J., of Cambridge, Mass. Convergence criteria for double Fourier series 29
GRAVES, L. M., of Chicago, Ill. A transformation of the problem of Lagrange in the calculus of variations 675
GRIFFIN, M., of Durham, N. C. Invariants of Pfaffian systems . . . 929
GROVE, V. G., of East Lansing, Mich. Contributions to the theory of transformations of nets in a space S_n 683
HESTENES, M. R., of Chicago, Ill. Sufficient conditions for the general problem of Mayer with variable end points 479

PAGE

HESTENES, M. R., and BLISS, G. A., of Chicago, Ill. Sufficient conditions for a problem of Mayer in the calculus of variations 305

HOLLCROFT, T. R., of Aurora, N. Y. The general web of algebraic surfaces of order n and the involution defined by it 855

HUNTINGTON, E. V., of Cambridge, Mass. New sets of independent postulates for the algebra of logic, with special reference to Whitehead and Russell's *Principia Mathematica* 274

——— Boolean algebra. A correction 557

——— A second correction 971

JEFFERY, R. L., of Wolfville, Nova Scotia. Sets of k-extent in n-dimensional space 629

LATIMER, C. G., of Lexington, Ky. On the class number of a cyclic field . 411

LEFSCHETZ, S., and WHITEHEAD, J. H. C., of Princeton, N. J. On analytical complexes 510

LEV, J., of Ithaca, N. Y. Effects of linear transformations on the divergence of bounded sequences and functions 888

LEWIS, D. C., of Cambridge, Mass. Infinite systems of ordinary differential equations with applications to certain second-order partial differential equations 792

McCOY, N. H., of Northampton, Mass. On the resultant of a system of forms homogeneous in each of several sets of variables 215

McSHANE, E. J., of Chicago, Ill. Parametrizations of saddle surfaces, with application to the problem of Plateau 716

MANNING, W. A., of Palo Alto, Calif. The degree and class of multiply transitive groups, III. 585

MILLER, G. A., of Urbana, Ill. Groups in which every operator has at most a prime number of conjugates 897

MONTGOMERY, D., of Iowa City, Iowa. Sections of point sets 915

MYERS, S. B., of Cambridge, Mass. Sufficient conditions in the problem of the calculus of variations in n-space in parametric form and under general end conditions. 746

ORE, O., of New Haven, Conn. On a special class of polynomials . . . 559

PALEY, R. E. A. C., of Cambridge, Mass. A special integral function . . 709

PALEY, R. E. A. C., and WIENER, N., of Cambridge, Mass. Notes on the theory and application of Fourier transforms. I–II 348

——— Notes on the theory and application of Fourier transforms. III, IV, V, VI, VII 761

PALL, G., of Montreal, Canada. The structure of the number of representations function in a binary quadratic form 491

RADÓ, T., of Columbus, Ohio. An iterative process in the problem of Plateau 869

RADÓ, T., and BECKENBACH, E. F., of Columbus, Ohio. Subharmonic functions and minimal surfaces 648

PAGE

——— Subharmonic functions and surfaces of negative curvature . . . 662

ROTH, W. E., of Milwaukee, Wis. On the equation $P(A, X)=0$ in matrices 689

SAKS, S., of Warsaw, Poland. On some functionals 549

——— Addition to the note on some functionals 965

SCHERBERG, M. G., of Minneapolis, Minn. The degree of convergence of a series of Bessel functions 172

SCHOENBERG, I. J., of Chicago, Ill. On finite-rowed systems of linear inequalities in infinitely many variables. II 452

SHEFFER, I. M., of State College, Pa. On the properties of polynomials satisfying a linear differential equation: Part I. 184

SHERMAN, J., of Philadelphia, Pa. On the numerators of the convergents of the Stieltjes continued fractions 64

SINGER, J., of Princeton, N. J. Three-dimensional manifolds and their Heegaard diagrams 88

SINKOV, A., of Washington, D. C. Families of groups generated by two operators of the same order 372

SNYDER, V., of Ithaca, N. Y. On a series of involutorial Cremona transformations of space defined by a pencil of ruled surfaces 341

SULLIVAN, M. M., of Cambridge, Mass. On the derivatives of Newtonian and logarithmic potentials near the acting masses 137

THOMAS, J. M., of Durham, N. C. Pfaffian systems of species one. . . . 356

WARD, M., of Pasadena, Calif. The cancellation law in the theory of congruences to a double modulus 254

——— The arithmetical theory of linear recurring series 600

WHITEHEAD, J. H. C., and LEFSCHETZ, S., of Princeton, N. J. On analytical complexes 510

WHITNEY, H., of Princeton, N. J. A characterization of the closed 2-cell . 261

WIENER, N., and PALEY, R. E. A. C., of Cambridge, Mass. Notes on the theory and application of Fourier transforms. I–II 348

——— Notes on the theory and application of Fourier transforms. III, IV, V, VI, VII 761

WIENER, N., of Cambridge, Mass., and YOUNG, R. C., of Cambridge, England. The total variation of $g(x+h)-g(x)$ 327

YOUNG, R. C., of Cambridge, England, and WIENER, N., of Cambridge, Mass. The total variation of $g(x+h)-g(x)$ 327

ZIPPIN, L., of Princeton, N. J. Correction to a paper on the Moore-Kline problem 972

SYSTEMS OF LINEAR DIFFERENCE EQUATIONS AND EXPANSIONS IN SERIES OF EXPONENTIAL FUNCTIONS*

BY
R. D. CARMICHAEL

Introduction. The principal purpose of the first part of this paper is to prove (§1.9) that the system (1.1) of linear non-homogeneous generalized difference equations has solutions $g_k(x)$, $k=1, 2, \cdots, n$, which are integral functions provided that the independent terms $\phi_\nu(x)$ are themselves integral functions and provided that the system has a certain non-singular character defined in §1.3. In case the $\phi_\nu(x)$ are further restricted to be of exponential type (§1.5) then solutions of exponential type exist (§1.6) and indeed solutions of exponential type at most equal to q (called principal solutions) in case no $\phi_\nu(x)$ is of higher type than q and at least one of them is of precisely this type. A useful symbolic notation (§1.2) is effective in carrying out the argument.

In the second part of the paper we apply the results of the first part to the rather remarkable problem of the simultaneous expansion of n integral functions in composite power series, a problem which we have not seen treated elsewhere.

The third part of the paper is devoted to the theory of a class of remarkable expansions in series of exponential functions, generalizing the theory of Fourier series. Whereas the basic region of convergence of Fourier series is a segment of a straight line, these new series, apart from certain particular cases, have certain polygons in the complex plane as their basic regions of convergence. The vertices of these polygons play the rôle of the end points of the segments in the case of Fourier series, while the remaining points of the polygon play the rôle of interior points of the segments. Several extensions of the theory are briefly indicated (§3.4) and an application is made (§3.5) to the expansion of Bernoulli polynomials of higher order in series of exponential functions.

I. On a system of linear difference equations with constant coefficients

1.1. Formulation of the problem. We consider the problem of solving the system

* Presented to the Society, August 31, 1932; received by the editors June 27, 1932.

$$\sum_{j=1}^{n} c_{\nu j} g_j(x + a_{\nu j}) = \phi_\nu(x) \qquad (\nu = 1, 2, \cdots, n) \tag{1.1}$$

of functional equations (generalized linear difference equations with constant coefficients), where the functions $\phi_\nu(x)$, $\nu = 1, 2, \cdots, n$, are n given integral functions and the n functions $g_j(x)$, $j = 1, 2, \cdots, n$, are to be determined subject to the requirement that they shall be integral functions. In this system the coefficients $c_{\nu j}$ and the additive terms $a_{\nu j}$ in the arguments are given constants; in §1.3 we shall subject these constants to a certain negative condition in order to avoid exceptional cases in the theory of the system.

We shall sometimes subject the $\phi_\nu(x)$ to additional restrictions and in such cases we shall put like further restrictions upon the solutions $g_j(x)$, thus obtaining what may be called principal solutions of the given system.

The theory of system (1.1) contains that of the single equation

$$\sum_{k=1}^{\mu} \gamma_k F(x + \alpha_k) = G(x) \tag{1.2}$$

where $\alpha_1, \alpha_2, \cdots, \alpha_\mu$ are different constants, $\gamma_1, \gamma_2, \cdots, \gamma_\mu$ are constants different from 0, $G(x)$ is an integral function, and $F(x)$ is to be determined as an integral function. To see this it is sufficient to write $g_k(x) = F(x + \alpha_k)$ and to form the system

$$g_1(x - \alpha_1) - g_k(x - \alpha_k) = 0, \quad k = 2, 3, \cdots, \mu, \quad \sum_{k=1}^{\mu} \gamma_k g_k(x) = G(x).$$

This system is of the form (1.1). From a solution of this system we have a solution of equation (1.2); and vice versa.

In a similar way one may reduce the problem of solving a system generalizing (1.1) and (1.2) at the same time to the problem of solving a system of the same form as (1.1).

Special cases of the problem here set have been treated by various authors.*

1.2. **Introduction of symbolic operators.** We define the symbolic operator $E(a)$ by the relation

$$E(a) \cdot f(x) \equiv f(x + a). \tag{1.3}$$

A linear homogeneous combination of such operators will have the meaning indicated by the relation

* See, for instance, C. Guichard, Annales de l'Ecole Normale Supérieure, (3), vol. 4 (1887), pp. 361–380; A. Hurwitz, Acta Mathematica, vol. 20 (1897), pp. 285–312; S. Pincherle, Ibid., vol. 48 (1926), pp. 279–304 (first published in 1888); R. D. Carmichael, American Journal of Mathematics, vol. 35 (1913), pp. 163–182; E. Hilb, Mathematische Annalen, vol. 85 (1922), pp. 89–98.

$$\left\{\sum_{k=1}^{\mu}\gamma_k E(a_k)\right\}\cdot f(x) \equiv \sum_{k=1}^{\mu}\gamma_k f(x + a_k). \tag{1.4}$$

This will serve, in particular, to define the sum and the difference of two operators of the form $\alpha E(a)$ and $\beta E(b)$. The product of these two operators is defined by the formula

$$\alpha E(a)\cdot\beta E(b) = \alpha\beta E(a + b). \tag{1.5}$$

These definitions serve to give a unique meaning to any polynomial combination of operators of the form $E(a_k)$, the coefficients being constants. Such a polynomial in operators E may be written as a linear function of suitably defined operators E, as one sees by aid of (1.5). In particular, one may define such an operator by means of a symbolic determinant of the form

$$\Delta \equiv \left| c_{\nu j}E(a_{\nu j})\right| \tag{1.6}$$

whose element in νth row and jth column is $c_{\nu j}E(a_{\nu j})$, this being (by definition) the symbolic operator obtained by expanding the determinant formally as if its elements were ordinary algebraic quantities. The expanded determinant may be written as a linear homogeneous function of suitable operators E with constant coefficients.

Any polynomial combination of such operators E will be said to have the value zero when and only when the result of operating with it upon an arbitrary integral function gives the value zero identically. It is easily shown that such a polynomial in operators E is zero if and only if the function e^{xt} of x is reduced to zero for all t when operated upon by the named operator.

1.3. Symbolic form of (1.1); restriction on the system. Employing the symbolic operators introduced in §1.2 we write system (1.1) in the form

$$\sum_{j=1}^{n} c_{\nu j}E(a_{\nu j})\cdot g_j(x) = \phi_\nu(x) \qquad (\nu = 1, 2, \cdots, n). \tag{1.7}$$

The determinant Δ in (1.6) will be called the symbolic determinant of system (1.7). This determinant will be called singular when it has the value zero; otherwise it will be called non-singular.

We shall treat system (1.1) or (1.7) only in the case when its determinant is non-singular. In that case we shall say that the system is non-singular. For the treatment of the excluded exceptional case the methods required are quite different from those here employed.

We shall use (without further definition) the terms customarily employed in the theory of determinants.

1.4. **Separation of variables.** Let $A_{\nu j}$ denote the cofactor of the element in the νth row and jth column of Δ. Then $A_{\nu j}$ is a polynomial in operators E with constant coefficients. Moreover, we have

$$(1.8)\qquad \sum_{j=1}^{n} c_{\nu j}E(a_{\nu j})A_{\nu j} = \delta_{\nu k}\Delta, \quad \sum_{\nu=1}^{n} c_{\nu j}E(a_{\nu j})A_{\nu k} = \delta_{jk}\Delta,$$

where $\delta_{\nu k}$ is 1 or 0 according as $\nu = k$ or $\nu \neq k$.

Multiplying the νth equation in (1.7) by the operator $A_{\nu k}$, summing as to ν from 1 to n, interchanging the order of summations in the first member of the resulting equation and simplifying by aid of the second equation in (1.8), we have

$$(1.9)\qquad \Delta g_k(x) = \sum_{\nu=1}^{n} A_{\nu k}\phi_\nu(x) \qquad (k = 1, 2, \cdots, n).$$

In these n equations the unknown functions $g_k(x)$ appear singly. Any solution of (1.7) must satisfy (1.9).

The operator Δ may be written in the form

$$(1.10)\qquad \Delta \equiv \sum_{k=0}^{\sigma} c_k E(a_k)$$

where the c_k are constants different from zero and the a_k are different constants (this form being surely possible since Δ is non-singular). The value of σ depends on n and the constants $c_{\nu j}$ and $a_{\nu j}$; it is never greater than $n!-1$.

If $\sigma = 0$ the required inverse operator Δ^{-1} is $E(-a_0)/c_0$. In this case a complete solution is readily obtained and the problem is trivial.

If $\sigma > 0$, as we shall henceforth assume it to be, we may write each equation (1.9) in the form

$$(1.11)\qquad c_0 g(x + a_0) + c_1 g(x + a_1) + \cdots + c_\sigma g(x + a_\sigma) = \phi(x),$$

where $\phi(x)$ is a given function and $g(x)$ is to be determined. From the solution of such an equation as this we shall pass to the solution of system (1.7).

We shall find it convenient to employ the function $h(t)$ defined by the equation

$$(1.12)\qquad h(t) = e^{-xt}\cdot\Delta\cdot e^{xt}.$$

Since the given system is non-singular it follows that $h(t)$ is not identically zero.

We have seen that every solution of (1.7) is a solution of (1.9). But the converse does not hold, as we shall now show. If $g_k(x)$, $k = 1, 2, \cdots, n$, is any solution of (1.9) then its general solution is $g_k(x) + p_k(x)$, $k = 1, 2, \cdots, n$,

where the functions $p_k(x)$ are arbitrary functions satisfying the equation $\Delta p_k(x)=0$. Now consider the system

$$g_1(x+1)+g_2(x)+g_3(x+2)=\phi_1(x),\ g_1(x+2)+g_2(x)+g_3(x+1)=\phi_2(x),$$
$$g_3(x)=\phi_3(x).$$

Here we have $\Delta=E(1)-E(2)\neq 0$, whereas the function $g_3(x)$ is uniquely determined by the last equation in the system.

From this example it follows that it is necessary to obtain an appropriate solution of (1.9) in order to have a solution of (1.7). The problem falls naturally into two cases; the following section prepares the way for this separation of cases.

1.5. **Functions of exponential type.** If $f(x)$ is an analytic function which is regular at x_0 and x_1, then it is easily shown that

$$\limsup_{\nu=\infty} |f^{(\nu)}(x_0)|^{1/\nu}=\limsup_{\nu=\infty} |f^{(\nu)}(x_1)|^{1/\nu},$$

where the superscripts denote derivatives with respect to x. If these superior limits have the finite value q ($q\geqq 0$) then $f(x)$ is an integral function; in such a case we shall say that $f(x)$ is of exponential type q, this terminology being justified by the following theorem,* stated here without proof:

THEOREM 1.1. *A necessary and sufficient condition that the function $f(x)$ shall be of exponential type q is* (1) *that numbers τ shall exist for which it is true that for every positive number ϵ there exists a quantity M, depending on ϵ and τ in general but independent of x, such that for all* (*finite*) *values of x, we have*

$$|f(x)|<Me^{(\tau+\epsilon)|x|}$$

and (2) *that q shall be the least possible value for such numbers τ. Moreover, when $f(x)$ is of exponential type q, we have*

$$|f^{(\nu)}(x)|<M(q+\epsilon)^{\nu}e^{(q+\epsilon)|x|} \qquad (\nu=0,1,2,\cdots),$$

where M is independent of x and ν.

1.6. **Case when the $\phi_\nu(x)$ are of exponential type.** We first carry out the solution of (1.7) for the case when the known functions $\phi_\nu(x)$ are of exponential type not exceeding q. Taking their power series expansions in the form

$$\phi_\nu(x)=\sum_{k=0}^{\infty}s_{\nu k}x^k/k! \qquad (\nu=1,2,\cdots,n), \tag{1.13}$$

* For a proof of this theorem and for further properties of functions of exponential type q, together with references to the literature, see a forthcoming paper of mine in Annals of Mathematics.

we introduce the functions $\psi_\nu(t)$ by means of the expansions

$$\psi_\nu(t) = \frac{s_{\nu 0}}{t} + \frac{s_{\nu 1}}{t^2} + \frac{s_{\nu 2}}{t^3} + \cdots \qquad (\nu = 1, 2, \cdots, n). \tag{1.14}$$

Then the series in (1.14) all converge if $|t| > q$. Let r be a positive number exceeding q such that the circle C_r of radius r about 0 as a center passes through no zero of the function $h(t)$ defined in (1.12). (This negative condition on r is first needed in the next paragraph.) Then we have

$$\phi_\nu(x) = \frac{1}{2\pi i}\int_{C_r} e^{xt}\psi_\nu(t)dt \qquad (\nu = 1, 2, \cdots, n), \tag{1.15}$$

as one sees by using expansions (1.14) and integrating term by term in (1.15).

We employ the operator $A_{\nu j}$ with the meaning given in §1.4. By $A_{\nu j}e^{xt}$ we mean the result of operating with $A_{\nu j}$ on e^{xt} considered as a function of x. Now write

$$g_j(x) = \frac{1}{2\pi i}\int_{C_r} \sum_{k=1}^{n}(A_{kj}e^{xt})\psi_k(t)\frac{dt}{h(t)} \quad (j = 1, 2, \cdots, n), \tag{1.16}$$

this being suggested by the problem of solving (1.9) by the method employed by Pincherle (loc. cit.) for a similar equation. Now substitute in the first member of (1.7) the functions $g_j(x)$ so defined and simplify by aid of (1.8) and (1.15); thus we have

$$\begin{aligned}\sum_{j=1}^{n} c_{\nu j}E(a_{\nu j})\cdot g_j(x) &= \frac{1}{2\pi i}\int_{C_r}\sum_{k=1}^{n}\left[\sum_{j=1}^{n} c_{\nu j}E(a_{\nu j})A_{kj}\cdot e^{xt}\right]\psi_k(t)\frac{dt}{h(t)} \\ &= \frac{1}{2\pi i}\int_{C_r}\sum_{k=1}^{n}\delta_{\nu k}(\Delta e^{xt})\psi_k(t)\frac{dt}{h(t)} \\ &= \frac{1}{2\pi i}\int_{C_r} e^{xt}\psi_\nu(t)dt = \phi_\nu(x).\end{aligned}$$

Therefore the functions $g_j(x)$ defined by (1.16) afford a solution of (1.7).

Equation (1.16) may be written in the form

$$g_j(x) = \frac{1}{2\pi i}\int_{C_r} e^{xt}\sum_{k=1}^{n}[A_{kj}e^{xt}]_{x=0}\psi_k(t)\,\frac{dt}{h(t)}\,.$$

Thence it follows that a constant M exists such that

$$|\,g_j^{(\nu)}(0)\,| < Mr\cdot r^\nu.$$

Therefore $g_j(x)$ is an integral function of exponential type not exceeding r.

Suppose next that r is so chosen that there is no zero of $h(t)$ in the interior of the circular ring bounded by C_r and the circle $|t|=q$. Then $g_j(x)$ remains unaltered as r decreases towards q remaining greater than q. Therefore, in this case, the function $g_j(x)$ is of exponential type not greater than q. Furthermore, if in this case at least one of the functions $\phi_\nu(x)$, $\nu=1, 2, \cdots, n$, is of exponential type q (none being of higher type), then at least one of the functions $g_j(x)$, $j=1, 2, \cdots, n$, is of exponential type q and none is of higher type.

If the functions $\phi_\nu(x)$ are of exponential type q or less and at least one of them is of type q then a solution of (1.7) will be called a *principal solution* if no function in it is of exponential type exceeding q. We have just shown the existence of such principal solutions. To determine all principal solutions we have to find all solutions, of exponential type not exceeding q, of the homogeneous system corresponding to (1.7). This problem is left for a later investigation.

The main result in this section may be stated as in the following theorem:

THEOREM 1.2. *When the $\phi_\nu(x)$ are functions of exponential type not greater than q and one at least of them is of type q, then the non-singular system* (1.1) *or* (1.7) *admits as a principal solution the functions $g_j(x)$ defined by* (1.16) *for $r=q+\epsilon$, where ϵ is a small positive quantity such that $h(t)$ has no zero in the ring bounded by C_r and the circle $|t|=q$.*

1.7. **Lemmas concerning exponential sums.** Equations (1.9) are of the form (1.11). Replacing x in (1.11) by $x-a_0$ we have another equation of the same form in which $a_0=0$. Hence there is no loss of generality in taking $a_0=0$; and this we do. Then the function $h(t)$ in (1.12) has the form

$$h(t) = c_0 + c_1e^{a_1t} + c_2e^{a_2t} + \cdots + c_\sigma e^{a_\sigma t}. \tag{1.17}$$

In preparation for the treatment of the case when the functions $\phi_\nu(x)$ are general integral functions we state certain lemmas concerning the function $h(t)$.

We shall first determine certain infinite regions in which $h(t)$ is free of zeros and in fact is bounded away from zero. Separating t and the a_k into real and imaginary parts we write

$$t = u + iv, \quad a_k = \alpha_k + i\beta_k \qquad (k = 1, 2, \cdots, \sigma).$$

Let $l_{\lambda\mu}$ denote the line

$$l_{\lambda\mu}: \qquad R(a_\lambda t) = R(a_\mu t), \ \lambda \neq \mu,$$

where $R(z)$ is the real part of z. Then $l_{\lambda\mu}$ and $l_{\mu\lambda}$ denote the same line. Moreover $l_{\lambda\mu}$ and $l_{\rho\tau}$ coincide if $a_\lambda-a_\mu=c(a_\rho-a_\tau)$ where c is a real number; otherwise they do not coincide. Let s denote the number of distinct lines in the set $l_{\lambda\mu}$,

Since each of these s lines passes through zero they divide the plane into $2s$ sectors such that no point of any one of these lines is in the interior of any such sector.

Let S_1 be any one of these $2s$ sectors and let t_1 be any given interior point of S_1. Then no two of the quantities $R(a_k t_1)$, $k=0, 1, \cdots, \sigma$, are equal. Let them be arranged in order of descending magnitude, thus:

$$R(a_{k_0}t_1) > R(a_{k_1}t_1) > \cdots > R(a_{k_\sigma}t_1).$$

Then if t_1 varies continuously over the interior of S_1 this continued inequality will be preserved, since each member varies continuously and no two become equal for an interior point of S_1. One and just one term of this continued inequality is zero for an interior point t_1 of S_1. Hence the first term is not negative.

In the sector S_1 take a point P which is at a distance δ from each of the bounding rays of the sector, where δ is a positive quantity whose value is to be assigned later. From P draw rays to infinity in S_1 and parallel to the bounding rays of S_1, thus forming a new sector S interior to S_1.

Let (u, v) be any point in S. Then the distance from (u, v) to the line $l_{k_0k_1}$ is the positive quantity

$$\frac{(\alpha_{k_0} - \alpha_{k_1})u - (\beta_{k_0} - \beta_{k_1})v}{\left\{(\alpha_{k_0} - \alpha_{k_1})^2 + (\beta_{k_0} - \beta_{k_1})^2\right\}^{1/2}}.$$

But this distance is not less than δ, whether $l_{k_0k_1}$ is or is not a bounding line of S_1. Therefore if t denotes the point (u, v) we have

$$\begin{aligned} R(a_{k_0}t) - R(a_{k_1}t) &= (\alpha_{k_0} - \alpha_{k_1})u - (\beta_{k_0} - \beta_{k_1})v \\ &\geqq \delta\left\{(\alpha_{k_0} - \alpha_{k_1})^2 + (\beta_{k_0} - \beta_{k_1})^2\right\}^{1/2}. \end{aligned}$$

Now let η be a fixed quantity such that

$$|c_\tau|(1 + e^{-\eta}) - (|c_0| + |c_1| + \cdots + |c_\sigma|)e^{-\eta} > 0 \qquad (\tau = 0, 1, \cdots, \sigma).$$

Let m be the least value attained by the left member as τ varies over the set $0, 1, \cdots, \sigma$. Determine δ so that

$$\delta\left\{(\alpha_\lambda - \alpha_\mu)^2 + (\beta_\lambda - \beta_\mu)^2\right\}^{1/2} \geqq \eta, \quad \lambda \neq \mu,$$

for every pair of different numbers λ and μ from the set $0, 1, \cdots, \sigma$. Then $R(a_{k_1}t) - R(a_{k_0}t) \leqq -\eta$ for all t in S. Hence $R(a_k t) - R(a_{k_0}t) \leqq -\eta$ for all t in S and for all k in the set $0, 1, \cdots, \sigma$ except $k=k_0$.

From these inequalities and the fact that $R(a_{k_0}t) \geqq 0$ in S it follows readily that for all t in S we have

$$|h(t)| \geqq m > 0. \tag{1.18}$$

This inequality is independent of the particular sector S; hence it holds for all sectors S formed (in the way indicated) by aid of a δ satisfying the named condition. We therefore have the following lemma:

LEMMA 1.1. *In the sectors S formed as indicated the function $h(t)$ satisfies inequality* (1.18).

When the sectors S are cut out of the plane there is left a sort of infinite star in which lie all the zeros of $h(t)$ and in fact all the points t for which $|h(t)| < m$. Thus we see that $h(t)$ is bounded away from zero in the distant part of the plane except possibly for certain regions in the star-arms remaining after removing the sectors S. We next consider the problem of bounding $h(t)$ away from zero in certain parts of these star-arms.

By a rotation of the t-plane, obtained by replacing t by $e^{i\theta}t$ where θ is real, any particular bounding ray of any sector S_1 may be transformed to the positive part of the u-axis. Since this transformation leaves invariant the sort of result we are to establish we may (and we shall) temporarily suppose that this transformation has already been carried out; for convenience we retain the original notation. The star-arm to be considered will then lie along the positive real axis; we denote it by A. Then the real axis is a line $l_{\lambda\mu}$, and we have $\alpha_\lambda = \alpha_\mu$ while $\beta_\lambda \neq \beta_\mu$.

The maximum α_k is positive or zero, since $\alpha_0 = 0$. If the maximum value α of the α_k is the value of just one of them, then as t becomes infinite in A, the function $|h(t)|$ becomes infinite or approaches a finite limit different from zero according as α is positive or zero. In this case $h(t)$ is bounded away from zero in the distant part of the star-arm.

In what remains we may therefore suppose that the maximum value α of the α_k is the value of two or more of them. Now in the star-arm A we have

$$|e^{-\alpha t}h(t)| \leqq |h(t)|,$$

the sign of equality holding when and only when $\alpha = 0$. But $e^{i\beta_k t}$ and its reciprocal are bounded in absolute value in A. It follows therefore that it is sufficient to treat only the special case in which $\alpha = 0$, as may be seen by replacing $h(t)$ by a suitable $e^{-(\alpha+i\beta_k)t}h(t)$. Therefore we take $\alpha = 0$. We temporarily choose the notation so that the values of k for which $\alpha_k = 0$ are $k = 0, 1, \cdots, \gamma-1$.

Write

$$h_1(t) = \sum_{k=0}^{\gamma-1} c_k e^{\beta_k(-v+iu)}. \tag{1.19}$$

Then $h(t) - h_1(t)$ approaches zero as t becomes infinite in A. It is therefore

sufficient to our purpose to determine suitable parts of A in which $h(t)$ is different from zero and $h_1(t)$ is bounded away from zero.

For this investigation we need the following classical lemma which we state without proof:

LEMMA 1.2. *If $b_1, b_2, \cdots, b_\nu$ is any set of real numbers, all different from zero, and if δ is any preassigned positive number, then there is an infinitude of positive integers m such that, for each such m, integers $k_1, k_2, \cdots, k_\nu$ exist such that*

$$|k_j b_j + m| \leqq \delta \qquad (j = 1, 2, \cdots, \nu). \tag{1.20}$$

If all such positive integers m are denoted by the symbols $m_1, m_2, \cdots$, with $m_j < m_{j+1}, j = 1, 2, \cdots$, then among the differences $m_{j+1} - m_j$ there is a greatest one.

Applying this lemma to the case when $b_j = 1/\beta_j$ and $\nu = \gamma - 1$, we have

$$|k_j + m\beta_j| \leqq \delta\beta_j \qquad (j = 1, 2, \cdots, \gamma - 1).$$

Thence it follows that for every preassigned positive ϵ there exists a δ such that we now have

$$|h_1(t + 2m\pi) - h_1(t)| \leqq \sum_{k=0}^{\gamma-1} |c_k e^{i\beta_k t}(e^{2\beta_k m\pi i} - 1)| < \epsilon$$

for all t in the star-arm A. Let R be a rectangle two sides of which are on the boundaries of A and let it be subject to the condition that $h_1(t)$ does not vanish in R. Let ϵ be such that $|h_1(t)| > 2\epsilon$ in R. Then

$$|h_1(t + 2m\pi)| > \epsilon$$

when t is in R and m is an integer admitted by the foregoing lemma.

We now return to the original form of $h(t)$ as given in (1.17). On each arm of the star associated with $h(t)$ we now take a rectangle R obtained from the foregoing one by reversing the rotation by which the corresponding arm is put in the special position employed in the preceding argument; or, we take any rectangle R on the arm and in which $h(t)$ does not vanish, in case the situation is such that the preceding argument reaches the goal before the introduction of Lemma 1.2 and the rectangle R. Then $h(t)$ is bounded away from 0 on R and on all congruent rectangles (except a finite number at most) similar to those in the preceding paragraph and containing the points $t + 2m\pi$ with t on R and m determined as in the lemma or m sufficiently large when the lemma is not needed.

A part of the foregoing results may be stated in the following lemma:*

LEMMA 1.3. *There exists a positive number ϵ such that $|h(t)| > \epsilon$ for all large t in sectors S and for all large t in rectangles R or rectangles obtained from them by the translations $t' = t + 2m\pi$ where t is in R and where m is an integer admitted by Lemma* 1.2 *for the star-arm in question or m is any sufficiently large integer in the cases where Lemma* 1.2 *is not employed in the argument.*

For use in integrations later to be performed let us define a set of contours $\Gamma_1, \Gamma_2, \cdots$, passing through no zero of $h(t)$, such that 0 is interior to Γ_1 while Γ_j is interior to Γ_{j+1} and such that for r greater than some preassigned number the distance from 0 to a point of Γ_r is not less than r and not greater than $r+\beta$ where β is a sufficiently large given positive number, each contour having the property that it consists of circular arcs (with 0 as center) in the sectors S and segments of the boundaries of the star-arms and straight line segments crossing these arms in the rectangles R or such rectangles congruent to them as are admitted by Lemma 1.3 and the preceding discussion. Then the length of Γ_r bears a bounded ratio to $2\pi r$.

From Lemma 1.3 we then have the following:

LEMMA 1.4. *There exists a positive number ϵ such that $|h(t)| > \epsilon$ for every t on every contour $\Gamma_1, \Gamma_2, \cdots$.*

From the distribution of the numbers m_j as described in Lemma 1.2, it follows that the contours $\Gamma_1, \Gamma_2, \cdots$ may be further restricted so that there exists a number p such that no more than p of the contours cross a given rectangle congruent to a given rectangle R in accordance with Lemma 1.3.

1.8. **Solution of equation** (1.11). In equation (1.11) we take $a_0 = 0$, as we may do without loss of generality. We now propose to show that, when $\phi(x)$ is any given integral function, this equation has a solution $g(x)$ which is itself an integral function.

We denote by $G_n(x)$ the polynomial which satisfies the equation

$$c_0 G_n(x) + \sum_{k=1}^{\sigma} c_k G_n(x + a_k) = x^n$$

and is (sometimes more precisely) defined by the formula

$$G_n(x) = \frac{n!}{2\pi i} \int_C \frac{e^{xt}}{h(t)} \frac{dt}{t^{n+1}},$$

* For such results as those in Lemmas 1.3 and 1.4 see the address of R. E. Langer, Bulletin of the American Mathematical Society, vol. 37 (1931), pp. 213–239, and the papers there cited, especially those of J. D. Tamarkin.

where n is a positive integer or zero, $h(t)$ denotes the function defined in (1.17) and C is a contour inclosing the point 0 and no singularity of the integrand other than $t=0$.

Let Γ_r, where r is any positive integer, denote the contour represented by this symbol in the latter part of §1.7. Form the function

$$G_{n,r}(x) = \frac{n!}{2\pi i}\int_{\Gamma_r} \frac{e^{xt}}{h(t)} \frac{dt}{t^{n+1}}.$$

This function satisfies the equation

$$c_0 G_{n,r}(x) + \sum_{k=1}^{\sigma} G_{n,r}(x + a_k) = x^n.$$

Let x be now confined to any preassigned finite region T of the x-plane. Then we have

$$|G_{n,r}(x)| \leqq \frac{n!}{2\pi} M_1 \int_{\Gamma_r} \frac{|e^{xt}|\,|dt|}{|t|^{n+1}},$$

where M_1 is a dominant of $|1/h(t)|$ for all t on all contours Γ_r, the existence of this dominant being assured by Lemma 1.4. From the character of the contours Γ_r, as described in the latter part of §1.7, we now see that a constant M (independent of x and n and r) exists such that for all x in T we have

$$|G_{n,r}(x)| < M \cdot n! \cdot \rho^{r+\beta} \cdot r^{-n}, \tag{1.21}$$

where ρ is such that $\rho > e^{|x|}$ for all x in T.

Write the power series expansion of $\phi(x)$ in the form

$$\phi(x) = \sum_{\nu=0}^{\infty} \lambda_\nu x^\nu. \tag{1.22}$$

Form the function $g(x)$,

$$g(x) = \sum_{\nu=0}^{\infty} \lambda_\nu G_{\nu,\nu}(x). \tag{1.23}$$

Then the $(\nu+1)$th term of the series here written is, in the region T and for sufficiently large values of ν, less in absolute value than the quantity

$$M \cdot \nu! \nu^{-\nu} \rho^{\nu+\beta} |\lambda_\nu|.$$

As ν becomes infinite the superior limit of the νth root of this quantity is zero since $|\lambda_\nu|^{1/\nu}$ has the superior limit zero owing to the fact that $\phi(x)$ is an integral function. Therefore the series in (1.23) converges absolutely and uni-

formly in any whatever preassigned finite region T. Since each term of this series is analytic throughout the finite plane it follows that $g(x)$ is itself an integral function.

It is readily verified by a direct substitution and a use of the named properties of $G_{\nu,\nu}(x)$ that this function $g(x)$ satisfies equation (1.11) with $a_0=0$.

We are thus led to the following theorem:

THEOREM 1.3. *If $\phi(x)$ denotes the integral function defined in* (1.22) *then the series*

$$\sum_{\nu=0}^{\infty}\lambda_\nu G_{\nu,r}(x) \tag{1.24}$$

is for suitable values of r absolutely and uniformly convergent in every finite region of the complex plane (the value $r=\nu$ being always suitable) and defines a sum function $g(x)$ which is an integral function of x and satisfies equation (1.11) *with $a_0=0$.*

If $\phi(x)$ is further restricted to be of exponential type q then it is easy to show (compare §1.6) that r may be given a sufficiently large fixed value (independent of ν) in series (1.24) to insure convergence of the character indicated in the theorem. In fact, Γ_r may be replaced by the circle C_r of §1.6. Then the resulting solution $g(x)$ of (1.11) is of exponential type not exceeding r where r is the radius of the circle C_r. By taking r sufficiently small it may be brought about that the resulting solution $g(x)$ is of exponential type q; but there is no solution $g(x)$ of lower type than q. When $\phi(x)$ is of exponential type q a solution $g(x)$ of (1.11) of exponential type q may be called a principal solution of that equation.

1.9. **The general case of (1.1) when the $\phi_\nu(x)$ are integral functions.** In treating this case it is convenient to set forth first a particular solution of system (1.9). Again and without loss of generality we take $a_0=0$.

Form the functions

$$g_k(x) = \frac{1}{2\pi i}\sum_{j=0}^{\infty}\sum_{\nu=1}^{n}\int_\Gamma s_{\nu j}(e^{-xt}A_{\nu k}e^{xt})\frac{e^{xt}}{h(t)}\frac{dt}{t^{j+1}} \quad (k=1,2,\cdots,n), \tag{1.25}$$

where the coefficients $s_{\nu j}$ are those appearing in (1.13). Now the expression in parenthesis under the integral sign is a function of t. If one utilizes the form of this function of t then by means of an easy modification of the argument employed in §1.8 one may show that the series in (1.25) converge absolutely and uniformly in every preassigned finite region T of the x-plane and that they define integral functions $g_k(x)$.

These integral functions may also be written in the form

$$(1.26)\qquad g_k(x) = \frac{1}{2\pi i} \sum_{j=0}^{\infty} \int_{\Gamma_j} \left(\sum_{\nu=1}^{n} s_{\nu j} A_{\nu k} e^{xt}\right) \frac{dt}{t^{j+1} h(t)} \qquad (k = 1, 2, \cdots, n),$$

the series having the same properties of convergence as before indicated.

Now by aid of (1.12) we have

$$\begin{aligned}\Delta g_k(x) &= \frac{1}{2\pi i} \sum_{j=0}^{\infty} \int_{\Gamma_j} \sum_{\nu=1}^{n} s_{\nu j} A_{\nu k} e^{xt} \frac{dt}{t^{j+1}} \\ &= \sum_{\nu=1}^{n} A_{\nu k} \sum_{j=0}^{\infty} \frac{s_{\nu j}}{2\pi i} \int_{\Gamma_j} e^{xt} \frac{dt}{t^{j+1}} \\ &= \sum_{\nu=1}^{n} A_{\nu k} \sum_{j=0}^{\infty} s_{\nu j} x^j / j! = \sum_{\nu=1}^{n} A_{\nu k} \phi_\nu(x),\end{aligned}$$

the last member being obtained from (1.13). Hence the functions $g_k(x)$ in (1.26) afford a solution of (1.9) with $a_0 = 0$.

That these same functions also afford a solution of (1.7) will next be proved. For this purpose substitute these functions $g_k(x)$ in the first member of (1.7) after replacing ν by μ. Simplifying the result by aid of equation (1.8) and other preceding formulas we have

$$\begin{aligned}\sum_{k=1}^{n} c_{\mu k} E(a_{\mu k}) g_k(x) &= \frac{1}{2\pi i} \sum_{j=0}^{\infty} \int_{\Gamma_j} \sum_{\nu=1}^{n} s_{\nu j} \left[\sum_{k=1}^{n} c_{\mu k} E(a_{\mu k}) A_{\nu k}\right] e^{xt} \frac{dt}{t^{j+1} h(t)} \\ &= \frac{1}{2\pi i} \sum_{j=0}^{\infty} s_{\mu j} \int_{\Gamma_j} \frac{\Delta e^{xt} dt}{t^{j+1} h(t)} \\ &= \frac{1}{2\pi i} \sum_{j=0}^{\infty} s_{\mu j} \int_{\Gamma_j} \frac{e^{xt} dt}{t^{j+1}} = \phi_\mu(x).\end{aligned}$$

Hence system (1.7) is satisfied by these functions $g_k(x)$.

Thus we have the following theorem:

THEOREM 1.4. *When the functions $\phi_\nu(x)$ in the non-singular system* (1.7) *are given integral functions and when the constant a_0 in* (1.10) *has the value* 0 *the system has a solution $g_k(x)$, $k = 1, 2, \cdots, n$, consisting of integral functions defined by equations* (1.26), *and the series in these equations converge absolutely and uniformly in every preassigned finite region T of the x-plane.*

From this theorem it follows that every non-singular system (1.1) has a solution consisting of integral functions whenever the given functions $\phi_\nu(x)$ are themselves integral. The more special case in which the $\phi_\nu(x)$ are of exponential type has already been treated in §1.6.

II. Simultaneous expansions of integral functions in composite power series

2.1. **Formulation of the problem.** For $n>1$ we consider the question of expanding n integral functions $f_1(x), f_2(x), \cdots, f_n(x)$ simultaneously in composite power series, that is, we consider the problem of representing these functions in the form

$$f_\nu(x) = \sum_{k=0}^{\infty}\sum_{j=1}^{n} c_{jk}(x - a_{\nu j})^k \qquad (\nu = 1, 2, \cdots, n), \tag{2.1}$$

where the coefficients c_{jk} are to be independent of both x and ν. We impose the further condition on the coefficients c_{jk} that they shall be such that the series in the equations

$$g_j(x) = \sum_{k=0}^{\infty} c_{jk}x^k \qquad (j = 1, 2, \cdots, n) \tag{2.2}$$

shall converge for all finite values of x; then the sum functions $g_j(x)$ defined by them will be integral functions. These conditions on the c_{jk} are equivalent to the conditions that the quantities $|c_{jk}|^{1/k}, j=1, 2, \cdots, n$, shall all have the limit zero as k becomes infinite. Furthermore we subject the given constants $a_{\nu j}$ to the condition that the determinant $\Delta(t)$ whose element in νth row and jth column is $\exp(-a_{\nu j}t)$ shall not be identically zero as a function of t. In the exceptional or singular case in which this condition on $\Delta(t)$ is not satisfied the general investigation will require methods different from those here employed; and the results will lack the simplicity and elegance which belong to the general case here treated.

For $n=1$ the problem evidently reduces to the classical problem of expansions in power series. We suppose throughout that $n>1$.

Under the conditions named we shall show that such simultaneous expansions always exist and indeed that they always exist subject to the further condition that the functions $g_j(x), j=1, 2, \cdots, n$, shall be of exponential type provided in the latter case that the functions $f_\nu(x), \nu=1, 2, \cdots, n$, are of exponential type.

If we employ the notation defined in (2.2) we may write (2.1) in the form

$$f_\nu(x) = \sum_{j=1}^{n} g_j(x - a_{\nu j}) \qquad (\nu = 1, 2, \cdots, n). \tag{2.3}$$

Integral solutions of this system evidently lead through (2.2) to the required expansions (2.1). The condition put on $\Delta(t)$ is just that which is required to make the results of the first part of this paper applicable to system (2.3) and hence to the expansion problem here set.

2.2. **Expansions in the case of general integral functions $f_\nu(x)$.** From Theorem 1.4 and the remark following it one concludes that system (2.3) has in this case integral solutions $g_j(x)$, $j=1, 2, \cdots, n$. Therefore we have the following theorem:

THEOREM 2.1. *If $f_1(x), f_2(x), \cdots, f_n(x)$ are any given integral functions and if the constants $a_{\nu j}$ are such that the determinant $\Delta(t)$ has the property described in the first paragraph of* §2.1, *then these functions $f_\nu(x)$ have simultaneous expansions of the form* (2.1) *where*

$$\lim_{k=\infty} |c_{jk}|^{1/k} = 0 \qquad (j = 1, 2, \cdots, n).$$

Formulas in §1.9 afford an effective means of obtaining suitable coefficients c_{jk} to be employed in the expansions (2.1). Only in exceptional cases is it true that these expansions are unique. The determination of the extent of arbitrary elements involved in the coefficients of the expansions depends on the (as yet undeveloped) theory of system (2.3) for the case when $f_\nu(x)\equiv 0$, $\nu=1, 2, \cdots, n$.

2.3. **Expansions when the $f_\nu(x)$ are of exponential type.** Applying Theorem 1.2 to system (2.3) in the case when the functions $f_\nu(x)$ are of exponential type and interpreting the results in terms of the expansions in (2.1), we have the following theorem:

THEOREM 2.2. *If the functions $f_1(x), f_2(x), \cdots, f_n(x)$ are of exponential type not exceeding q, one at least of them being precisely of type q, and if the constants $a_{\nu j}$ are such that the determinant $\Delta(t)$ has the property described in the first paragraph of* §2.1, *then the functions $f_\nu(x)$ have simultaneous expansions of the form* (2.1) *such that the associated functions $g_j(x)$ of* (2.2) *are of exponential type and indeed such that these functions $g_j(x)$ are of exponential type not exceeding q, one at least of them being precisely of type q.*

When the associated $g_j(x)$ are of exponential type not exceeding q we shall say that the series in (2.1) afford *principal expansions* of the functions $f_\nu(x)$.

Even with the strongest conditions imposed on the coefficients c_{jk} by the latter part of the foregoing theorem it is still true that the expansions (2.1) need not be unique. In all cases belonging to this section possible values of the coefficients c_{jk} are readily determined from the special case of equation (1.16) applicable here, as we show in the next paragraph; and these values may well vary in dependence upon the radius r of the circle C_r appearing in (1.16).

In connection with the expansions

$$f_\nu(x) = \sum_{k=0}^{\infty} \alpha_{\nu k} x^k/k! \qquad (\nu = 1, 2, \cdots, n),$$

form the functions

$$F_\nu(t) = \sum_{k=0}^{\infty} \frac{\alpha_{\nu k}}{t^{k+1}} \qquad (\nu = 1, 2, \cdots, n).$$

Let $\Delta_{\nu j}(t)$ be the cofactor of the element in the νth row and jth column of $\Delta(t)$. Then by aid of (1.16) it may readily be shown that suitable coefficients c_{jk} in (2.1) are the following:

$$c_{jk} = \frac{1}{2\pi i(k!)} \int_{C_r} \left(\sum_{\nu=1}^{n} \Delta_{\nu j}(t) F_\nu(t) \right) \frac{t^k dt}{\Delta(t)},$$

where $j=1, 2, \cdots, n$ and $k=0, 1, 2, \cdots$.

2.4. **The case $a_{\nu j}=a_{\nu\nu}$ for $\nu>j$.** In this case system (2.3) is equivalent to the system consisting of the first equation in (2.3) and the following $n-1$ equations:

$$(2.4)\quad f_{\nu-1}(x) - f_\nu(x) = \sum_{j=\nu}^{n} \{ g_j(x - a_{\nu-1,j}) - g_j(x - a_{\nu j}) \} \quad (\nu = 2, 3, \cdots, n).$$

In case $a_{n-1,n}=a_{nn}$ it is clear that we must have $f_{n-1}(x)=f_n(x)$ as a necessary condition for satisfying the system. In fact, it is easy to see that the functions $f_\nu(x)$ must satisfy one or more special restrictive conditions if one or more of the relations

$$(2.5)\qquad a_{\nu-1,\nu} - a_{\nu\nu} \neq 0 \qquad (\nu = 2, 3, \cdots, n)$$

fails to be satisfied. But if conditions (2.5) are all satisfied then we have an instance of the general theory already developed; we shall suppose that these conditions are satisfied. We assume that the given functions $f_1(x), \cdots, f_n(x)$ are all integral functions. We require that the functions $g_1(x), \cdots, g_n(x)$ shall be integral functions.

Taking $\nu=n$ in (2.4) we see that $g_n(x)$ is uniquely determined as an integral function except for an arbitrary additive periodic integral function of period $a_{n-1,n}-a_{nn}$. Taking $g_n(x)$ to be any integral function satisfying (2.4) for $\nu=n$ we may then determine $g_{n-1}(x)$ uniquely except for an additive integral function of period $a_{n-2,n-1}-a_{n-1,n-1}$. With $g_{n-1}(x)$ determined we proceed similarly to the determination of $g_{n-2}(x)$, and we continue thus until $g_2(x)$ is determined. Then the first equation in (2.3) uniquely determines $g_1(x)$. It appears, therefore, that in the present case one can determine com-

pletely the arbitrary elements in the solution of (2.3) subject to the named conditions. Hence all possible expansions (2.1) are completely determined for the present case.

If we further restrict the given functions $f_1(x), \cdots, f_n(x)$ to be of exponential type not greater than q we may likewise determine the functions $g_1(x), \cdots, g_n(x)$ so that they are of exponential type not greater than q and we may show precisely what is arbitrary in the determination of such functions subject to these conditions. These results may then be carried over to the corresponding case of the expansions (2.1).

There is one case of particular interest in which the expansions (2.1), when subject to the condition named in the preceding paragraph, are unique except for the trivial restriction that the constants c_{j0}, $j=1, 2, \cdots, n$, are not separately determined but only their sum is determined. This is the case in which the functions $f_1(x), \cdots, f_n(x)$ are of exponential type not greater than q while at the same time the relations

$$q \left| a_{\nu-1,\nu} - a_{\nu\nu} \right| < 2\pi \qquad (\nu = 2, 3, \cdots, n) \tag{2.6}$$

are all satisfied. For in this case each $g_j(x)$ is uniquely determined except for an additive constant. These conditions are obviously satisfied whenever inequalities (2.5) hold provided that $q=0$ and in particular provided that the functions $f_\nu(x)$ are polynomials.

2.5. **The case $n=2$.** For the case $n=2$ system (2.3) may be written in the form

$$\begin{aligned} f_1(x + a_{11}) &= g_1(x) + g_2(x + a_{11} - a_{12}), \\ f_2(x + a_{21}) &= g_1(x) + g_2(x + a_{21} - a_{22}). \end{aligned} \tag{2.7}$$

The exceptional case here is that in which $a_{11}-a_{12}=a_{21}-a_{22}$. When this condition is satisfied, the system can have a solution only when $f_1(x+a_{11}) = f_2(x+a_{21})$, as one sees from (2.7); and in this case it is clear that either of the integral functions $g_1(x)$ and $g_2(x)$ may be assigned at will and that the other is then uniquely determined: the case is therefore trivial.

When $a_{11}-a_{12} \neq a_{21}-a_{22}$ the case belongs to that treated in §2.4.

As an application of the case when $a_{11}=a$, $a_{12}=b$, $a_{21}=0=a_{22}$, where $a \neq b$, we see that an arbitrary integral function $f(x)$ may be expanded in the form

$$f(x) = \sum_{k=0}^{\infty} \{ c_k(x-a)^k + \gamma_k(x-b)^k \} \tag{2.8}$$

where the sums $c_k+\gamma_k$, $k=0, 1, 2, \cdots$, have any preassigned values subject to the condition that

$$\lim_{k=\infty} \left| c_k + \gamma_k \right|^{1/k}$$

shall exist and be equal to zero; and the parts of $f(x)$ represented by the component power series in $x-a$ and $x-b$ respectively, when these parts are themselves required to be integral functions, are unique except for an arbitrary integral periodic function of period $a-b$ to be added to one part and subtracted from the other.

Furthermore, if $f(x)$ is of exponential type not greater than q and if the parts of $f(x)$ represented by the component power series in $x-a$ and $x-b$ respectively are required to be of exponential type not greater than q, then there exists an expansion of the form (2.8) subject to the condition that

$$\limsup_{k=\infty} \left| (c_k + \gamma_k)/k! \right|^{1/k} \leqq q;$$

and the expansion is unique except for an arbitrary periodic function of period $a-b$ and of exponential type not greater than q, such periodic function to be added to one component part of $f(x)$ and subtracted from the other. If we add the further restriction that $q|a-b| < 2\pi$ then this periodic function reduces to a constant, so that the expansion (2.8) is then essentially unique.

2.6. **Generalizations.** From the fact established in §1.9 that the nonsingular system (1.1) always has integral solutions when the $\phi_\nu(x)$ are given integral functions it follows that any set $\phi_1(x), \cdots, \phi_n(x)$ of integral functions has simultaneous expansions in the form

$$(2.9) \qquad \phi_\nu(x) = \sum_{k=0}^{\infty} \sum_{j=1}^{n} \alpha_{kj} c_{\nu j} (x + a_{\nu j})^k \qquad (\nu = 1, 2, \cdots, n),$$

where the constants α_{kj} are independent of x and ν and where the component functions $g_j(x)$,

$$g_j(x) = \sum_{k=0}^{\infty} \alpha_{kj} x^k \qquad (j = 1, 2, \cdots, n),$$

are themselves integral functions. If the $\phi_\nu(x)$ are subject to the further condition that they shall be of exponential type not greater than q then the expansions (2.9) exist subject (as one sees from §1.6) to the condition that the component functions $g_j(x)$ shall also be of exponential type not greater than q. If furthermore at least one of the functions $\phi_\nu(x)$ is of precisely type q then one at least of the component functions $g_j(x)$ is of precisely type q.

These results are capable of extension by means of the generalizations indicated near the end of §1.1.

There is a special case arising from expansions (2.9) to which particular attention may be directed. Let $a_{1j} = -a_j$, $j = 1, 2, \cdots, n$, where $a_1, a_2, \cdots,$

a_n are different constants, and let the other $a_{\nu j}$ have the value 0. Let $c_{1j}=1$, $j=1, 2, \cdots, n$, while the other $c_{\nu j}$ are such that the matrix

$$\left\| \begin{matrix} c_{21} & c_{22} & \cdots & c_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{matrix} \right\|$$

is of rank $n-1$. Then the corresponding system (1.1) is non-singular. Consider the problem of expanding a given integral function $\phi(x)$ in the form

$$\phi(x) = \sum_{k=0}^{\infty} \sum_{j=1}^{n} \alpha_{kj}(x - a_j)^k. \tag{2.10}$$

Since $\phi(x)$ thus takes the place of $\phi_1(x)$ in (2.9) and since the remaining integral functions

$$\phi_2(x), \cdots, \phi_n(x)$$

in (2.9) may be assigned at will, it follows that an expansion of the form (2.10) exists (not necessarily unique) such that

$$\lim_{k=\infty} | \alpha_{kj} |^{1/k} = 0 \qquad (j = 1, 2, \cdots, n),$$

while the quantities

$$\beta_{\nu k} = \sum_{j=1}^{n} c_{\nu j}\alpha_{kj} \qquad (\nu = 2, \cdots, n; k = 0, 1, 2, \cdots)$$

may be assigned at will subject to the condition that

$$\lim_{k=\infty} | \beta_{\nu k} |^{1/k} = 0 \qquad (\nu = 2, \cdots, n).$$

This result affords an interesting generalization of the Cauchy-Taylor expansion of an integral function. Whether there exists a corresponding generalization for functions analytic in a finite region I have not sought to determine.

If $\phi(x)$ is further restricted to be of exponential type not greater than q then there exists an expansion of the form (2.10) (not necessarily unique) such that

$$\limsup_{k=\infty} | \alpha_{kj}/k! |^{1/k} \leqq q \qquad (j = 1, 2, \cdots, n),$$

while the quantities

$$\beta_{\nu k} \qquad (\nu=2, \cdots, n; k=0, 1, 2, \cdots)$$

may be assigned at will subject to the condition that

$$\limsup_{k=\infty} | \beta_{\nu k}/k! |^{1/k} \leqq q \qquad (\nu = 2, \cdots, n).$$

III. Expansions in series of exponential functions

3.1. Properties of exponential sums. Let us denote by $h(t)$ the function

$$h(t) = c_1e^{a_1t} + c_2e^{a_2t} + \cdots + c_ne^{a_nt}, \quad n > 1, \tag{3.1}$$

where $a_1, a_2, \cdots, a_n$ are different constants and $c_1, c_2, \cdots, c_n$ are constants different from zero. And let us consider the problem of bounding away from zero the function $e^{-xt}h(t)$ for suitable given values of x and for suitable ranges of t. The results are needed for our later investigation (§3.2) of certain contour integrals.

Let P be the smallest convex polygon, in the complex plane, containing the points $a_1, a_2, \cdots, a_n$; this polygon may in special cases reduce to a straight line segment. Let Q be the polygon* obtained by reflecting P through the real axis. For the sake of definiteness we suppose that the notation is so chosen that the vertices of P, taken in counter-clockwise order, are $a_1, a_2, \cdots, a_\nu$ $(\nu \leqq n)$ and that no a_j has its real part less than that of a_1. Moreover we suppose that the vertices are so taken that no three of these a's at the vertices lie on the same straight line. Let $l_1, l_2, \cdots, l_\nu$ be the rays normal to the sides of Q at their centers and drawn outward from this polygon; when Q reduces to a straight line it is to be understood that these rays are two in number and that they are drawn so that there is one in each direction from the middle point of the line. We take the notation so that $l_1, l_2, \cdots, l_\nu$ are in clockwise order and so that l_1 is the normal to the side joining the conjugates of a_1 and a_2.

Let a_j and a_k be two consecutive vertices of P and let l_μ be the normal to that side of Q which joins the corresponding vertices of Q. If $R(z)$ denotes the real part of z, then the line $R(a_jt) = R(a_kt)$ is parallel to the line l_μ. Let ρ be a positive number whose value is later to be conveniently restricted. On each side of each line $l_1, l_2, \cdots, l_\nu$ and at a distance ρ from it draw a ray in such a way that these rays will make a sort of infinite star similar to that considered in §1.7 and containing the rays $l_1, l_2, \cdots, l_\nu$ in the centers of its arms. These rays form certain sectors S, similar to those in §1.7 and containing no interior points of the named infinite star.

In order to have sectors exactly like those in §1.7 it is necessary to divide some of the sectors S into smaller sectors by excluding other strips; but this further division is to serve only a temporary purpose in the argument. It may be described as follows. Let $m_1, m_2, \cdots, m_\nu$ be rays from zero to infinity parallel to $l_1, l_2, \cdots, l_\nu$ respectively but such that m_k goes to infinity in a direction opposite to that of l_k. Some rays m_k may go to infinity in the same

* Such polygons as P and Q have been employed by Pólya, Mathematische Annalen, vol. 89 (1923), pp. 179–191.

direction as other rays l_j (and they will do so when Q has pairs of parallel sides); remove such rays m_k; if any rays m_σ remain after this removal, denote them by $m_\alpha, m_\beta, \cdots$. Along the rays $m_\alpha, m_\beta, \cdots$ remove strips of width 2ρ as in the case of the preceding paragraph. Then some sectors S are separated into two or more sectors (together with one or more strips). After all such separations are made, let S' be a symbol to denote the totality of sectors obtained, including undivided sectors S and the parts into which some sectors S have been separated.

From Lemma 1.1 it follows that ρ may be taken sufficiently large that $h(t)$, and hence $e^{-xt}h(t)$, shall have no zero in any sector S'. Moreover, from the same lemma it follows that ρ may be taken sufficiently large (and we so take it) that $e^{-xt}h(t)$ is bounded away from zero in the sectors S' when x is any one of the points $a_1, a_2, \cdots, a_n$. In fact, when x has any such value the function $e^{-xt}h(t)$ is a function meeting the conditions on $h(t)$ in §1.7 so that Lemmas 1.3 and 1.4 are also applicable to $e^{-xt}h(t)$ for such values of x.

Let a_j, a_k and a_l be any three consecutive vertices of P in counterclockwise order and let l_j and l_k be the rays perpendicular to the sides of Q with corresponding vertices. Let S_{jk} be the sector S lying between l_j and l_k. Suppose that t varies in S_{jk}. Let x be a fixed point in P. We have

$$\left| e^{-xt}h(t) \right| = \left| e^{(a_k - x)(t - \bar{a}_k)} \right| \cdot \left| e^{\bar{a}_k(a_k - x)} \right| \cdot \left| e^{-a_k t}h(t) \right|,$$

where $\bar{a}_k$ is the conjugate of a_k. The last factor in the second member is bounded away from zero for large t in the named sector, as we have already seen. The middle factor is a constant different from zero, since x is fixed. The argument of the exponent of the first factor lies between $-\frac{1}{2}\pi$ and $\frac{1}{2}\pi$ inclusive, as one may readily show graphically, if (as we do by taking ρ sufficiently large) we restrict the sector S_{jk} to lie in the sector formed by rays from $\bar{a}_k$ to infinity in the direction of the rays l_j and l_k: in establishing the named fact it is convenient temporarily to transform the points of the plane by adding $-\bar{a}_k$ to each value in it so that the representation of $\bar{a}_k$ becomes the point zero and then to begin from the plots of $\bar{x} - \bar{a}_k$ and $t - \bar{a}_k$. Thence it follows that $e^{-xt}h(t)$ is bounded away from zero in the named sector. Furthermore it follows from Lemma 1.3 that $e^{-xt}h(t)$ is bounded away from zero for all large t in rectangles congruent to the rectangles R in the way specified in that lemma, these rectangles R being chosen with reference to the function $e^{-xt}h(t)$.

Let us now further restrict x to lie in the interior of P. Then there exists a positive number ϵ such that

$$-\tfrac{1}{2}\pi + \epsilon \leqq \arg \left\{ (a_k - x)(t - \bar{a}_k) \right\} \leqq \tfrac{1}{2}\pi - \epsilon.$$

Thence it follows that the function $t^{-1}e^{-xt}h(t)$ is bounded away from zero as t becomes infinite in the named sector.

The same function is also bounded away from zero if x is on the boundary of P but not at a_k while t becomes infinite in the named sector in such a way as to remain outside of each of two parabolas with vertex at $\bar{a}_k$ and having the named rays from $\bar{a}_k$ parallel to l_j and l_k as their principal diameters.

It may now be observed that every strip along one of the rays $m_\alpha, m_\beta, \cdots$ lies (except for a finite part of it) entirely in a sector S and that it has a direction intermediate to the directions of the bounding rays of this sector S. Thence it follows also that such a strip (except for a finite part of it) lies entirely outside of the parabolas along the bounding rays of this sector S. Hence the strips along the rays $m_\alpha, m_\beta, \cdots$ may be removed and we thus return to the set of sectors S as defined in this section; and for the plane so divided we have the requisite character of $e^{-xt}h(t)$ or $t^{-1}e^{-xt}h(t)$ as a function bounded away from zero, in accordance with the paragraph next following.

Summing up these results we may state that $e^{-xt}h(t)$ is bounded away from zero for any given x in P and for all large t in all sectors S formed with sufficiently large ρ and in all rectangles congruent to rectangles R in accordance with Lemma 1.3; that $t^{-1}e^{-xt}h(t)$ is bounded away from zero for each interior point x of P and for all large t in all such sectors S; and that $t^{-1}e^{-xt}h(t)$ is bounded away from zero for each x on the boundary of P and not at a vertex of P and for all large t in all such sectors S and outside of all parabolas of the sort described for $\bar{a}_k$ in the previous paragraph, two such parabolas being formed at each vertex of Q.

3.2. **Properties of certain contour integrals.** Let $C_1, C_2, \cdots, C_s, \cdots$ be a set of different contours in the complex plane such that any given point on C_j is either interior to C_{j+1} or on C_{j+1} and such that for every s there exists an r such that the contour C_s is a contour Γ_r of the sort described in §1.7 and suitable to apply to $e^{-xt}h(t)$ for points x in P as the contours Γ_r apply to the function $h(t)$ of §1.7 and such that for every r there is an s such that C_s is a contour Γ_r.

Let $\psi(t)$ be any function of t which is analytic at infinity and vanishes there and let us write

$$(3.2)\qquad \psi(t) = \gamma_1/t + \gamma_2/t^2 + \gamma_3/t^3 + \cdots, |t| > q.$$

Let r be a fixed integer such that the contour C_r lies entirely within the region of convergence of the series in (3.2). Form the function $F_r(x)$,

$$(3.3)\qquad F_r(x) = \frac{1}{2\pi i}\int_{C_r} e^{xt}\{h(t)\}^{-1}\psi(t)dt.$$

Then $F_r(x)$ is a function of exponential type; and, in fact, it is such a function

as arises from the solution of equation (1.11) when $\phi(x)$ is a given function of exponential type, as one sees from Part I and especially from §1.6.

Let p be any positive integer and form the function $F_{r+p}(x)$ by changing r to $r+p$ in (3.3). We shall show that

$$\lim_{p=\infty} F_{r+p}(x) = 0 \tag{3.4}$$

when any one of the following conditions is satisfied:

(1) when x is in the interior of P;
(2) when x is on the boundary of P and is not a vertex of P;
(3) when x is a vertex of P provided in this case that $\gamma_1=0$.

It is convenient to carry out the proof first for the case when $\gamma_1=0$. Then a number M exists such that $|t^{-2}\psi(t)|<M$ on all the contours C_{r+p}. We let x be any point of P either in the interior or anywhere on the boundary. Then from the results at the end of §3.1 it follows that a constant M_1 exists such that $|e^{xt}\{h(t)\}^{-1}|<M_1$. Hence there is a constant M_2 such that in this case we have

$$|F_{r+p}(x)| < M_2 \int_{C_{r+p}} |t|^{-2}|dt|.$$

This implies the truth of (3.4) when $\gamma_1=0$ and x is anywhere in P.

With this result in hand we see that (3.4) will be established in the three cases (1), (2), (3) if we further prove its validity in cases (1) and (2) for the particular function $\psi(t)=1/t$, since we may then pass to the general case in an obvious manner.

In case (1) let us write

$$F_{r+p}(x) = \frac{1}{2\pi i}\left(\int_{S_{r+p}} + \int_{A_{r+p}}\right)e^{xt}\{h(t)\}^{-1}t^{-1}dt, \tag{3.5}$$

where S_{r+p} denotes the set of paths consisting of the parts of C_{r+p} which lie in the sectors S while A_{r+p} is the set of paths consisting of the remaining parts of C_{r+p}. Then on A_{r+p} the integrand has a dominant of the form $M/|t|$ while on S_{r+p} it has a dominant of the form $M/|t^2|$, as one sees from the results in the last paragraph of §3.1. Thence we conclude readily to the truth of (3.4) for the present case, since the total length of the parts A_{r+p} is bounded.

In case (2) we may use notationally the same equation (3.5) where we now understand that A_{r+p} denotes the set of paths consisting of the parts of C_{r+p} which lie in the parabolas described near the end of §3.1 while S_{r+p} consists of the remaining parts of C_{r+p}. The conclusion that (3.4) is valid in the present case is reached in the same way as in the preceding paragraph but by using

the additional fact that the total length of the parts A_{r+p} bears to the minimum distance d from zero to points of A_{r+p} a ratio which is infinitesimal as $r+p$ becomes infinite.

Thus the relation (3.4) is established for all points x of P except that when x is at a vertex of P we require that γ_1 shall have the value zero.

3.3. **Expansions in series of exponential functions.** Let $S_{r+p}(x)$ denote the negative of the sum of the residues of the function $e^{xt}\{h(t)\}^{-1}\psi(t)$ in the region bounded by the contours C_{r+p-1} and C_{r+p}. If the function has no singularity in this region we shall understand that $S_{r+p}(x)$ is identically zero. In all other cases $S_{r+p}(x)$ is a function of the form $ce^{\alpha x}$ or a sum of a finite number of such functions. We have

$$F_r(x) - F_{r+p}(x) = \sum_{k=1}^{p} S_{r+k}(x).$$

If we suppose that x is a point of P and in case $\gamma_1 \neq 0$ that it is not a vertex of P then relation (3.4) is applicable to the foregoing equation when p is allowed to become infinite and we have the following theorem:

THEOREM 3.1. *The function $F_r(x)$ defined in* (3.3) *has the expansion*

$$F_r(x) = \sum_{k=1}^{\infty} S_{r+k}(x) \tag{3.6}$$

in series of exponential functions, valid for all values of x in the polygon P, except that the vertices are to be excluded when $\gamma_1 \neq 0$.

In the special case when $h(t) = e^t - 1$ the series in (3.6) is a Fourier series. The polygon P in this case reduces to the interval (01) of the real axis, the end points of the interval serving as the vertices of the polygon. A further treatment of Fourier series from this point of view will appear in a forthcoming paper in Annals of Mathematics.

The foregoing theorem serves to expand in series (3.6) any whatever function that may be put in the form (3.3). If $h(0) \neq 0$ it is evident that any given polynomial in x may be put in the form $F_1(x)$ by taking C_1 to be a small circle about 0 as a center and by choosing $\psi(t)$ properly as a polynomial in $1/t$. The function $F_1(x)$+constant may also in other cases represent any whatever polynomial in x. Hence, in particular, all polynomials have expansions in the form (3.6), or in this form with an additive constant, valid in polygons P as indicated.

3.4. **A special class of the foregoing expansions.** We shall now examine the special case of the foregoing expansion theory in which the function $e^{xt}\{h(t)\}^{-1}$ has the form

$$\frac{e^{xt}}{(e^{\rho_1 t}-1)(e^{\rho_2 t}-1)\cdots(e^{\rho_n t}-1)}, \tag{3.7}$$

where $\rho_1, \rho_2, \cdots, \rho_n$ are n real or complex constants different from 0 and such that neither the sum nor the difference of two of them is zero.

It is convenient, for the sake of simplicity, to normalize the problem by means of certain elementary transformations. If ρ_k has a negative real part we may replace ρ_k by $-\rho_k$ by multiplying both numerator and denominator in (3.7) by $-e^{-\rho_k t}$ and so obtain (except for an irrelevant change in sign) a similar expression with x replaced by $x-\rho_k$; by a translation in the x-plane we may then replace $x-\rho_k$ by x. We suppose all such translations made so that we shall assume that the real part of each ρ_k is positive or zero. Then the further conditions on $\rho_1, \rho_2, \cdots, \rho_n$ are that they are different from each other and from zero. Then the point zero is on the boundary of the polygon P, introduced (§3.1) in the general case, and the greatest real value of a point in P is the sum of the real parts of $\rho_1, \rho_2, \cdots, \rho_n$. We suppose that the notation is so chosen that

$$-\tfrac{1}{2}\pi \leqq \arg \rho_1 \leqq \arg \rho_2 \leqq \cdots \leqq \arg \rho_n \leqq \tfrac{1}{2}\pi. \tag{3.8}$$

By means of a straight line join each point (except the last) in the set

$$\begin{gathered} 0, \rho_1, \rho_1+\rho_2, \rho_1+\rho_2+\rho_3, \cdots, \rho_1+\rho_2+\cdots+\rho_n, \\ \rho_2+\cdots+\rho_n, \cdots, \rho_{n-1}+\rho_n, \rho_n, 0 \end{gathered} \tag{3.9}$$

to the one which follows it, thus forming a convex polygon of an even number of sides and having its sides parallel in pairs. This is the polygon P, as one sees by examining points x in the sectors formed by adjacent sides. Then the points x in P are the points

$$x = \lambda_1\rho_1 + \lambda_2\rho_2 + \cdots + \lambda_n\rho_n \qquad (0 \leqq \lambda_k \leqq 1;\ k = 1, 2, \cdots, n), \tag{3.10}$$

as one sees by aid of the fact that each of these points lies in the strips each of which is bounded by two parallel sides of P and by showing that every point in P is a point x of the named form. The boundary of P is traced out in counter-clockwise order by starting with all λ's equal to zero, then letting λ_1 increase from 0 to 1, then λ_2 from 0 to 1, and so on to λ_n letting it increase from 0 to 1, then letting λ_1 decrease from 1 to 0, λ_2 from 1 to 0, and so on till λ_n decreases from 1 to 0.

For every point x in P the function (3.7) may be written in the form

$$\frac{e^{xt}}{(e^{\rho_1 t}-1)\cdots(e^{\rho_n t}-1)} = \frac{e^{\lambda_1\rho_1 t}}{e^{\rho_1 t}-1}\cdot\frac{e^{\lambda_2\rho_2 t}}{e^{\rho_2 t}-1}\cdot\cdots\cdot\frac{e^{\lambda_n\rho_n t}}{e^{\rho_n t}-1}, \tag{3.11}$$

where $0 \leqq \lambda_k \leqq 1$, $k=1, 2, \cdots, n$. For this special case the inequalities obtained in §3.1 may be derived in a very simple manner, as one may see by applying the methods of §§1.7 and 3.1 separately to each factor of the second member of (3.11) and simplifying the procedure in obvious ways for these special cases.

Moreover, when no two of the ρ_j have a real ratio, the contours $C_1, C_2, \cdots$ may be chosen so that C_k incloses just k zeros of $h(t)$ for $k=1, 2, \cdots$. Hence the terms $S_{r+k}(x)$ in (3.6) may all be taken in the form $ce^{\alpha x}$ so that we have to do with expansions of the form*

$$F(x) = \alpha_{00} + \sum_{m=1}^{\infty} \sum_{k=1}^{n} (\alpha_{km} e^{2m\pi i x/\rho_k} + \beta_{km} e^{-2m\pi i x/\rho_k}). \tag{3.12}$$

In what follows in this section we shall suppose that no two of the numbers ρ_k have a real ratio. Then no two terms in the series (3.12) involve the same exponential function.

With each of the functions

$$1,\ e^{2m\pi i x/\rho_k},\ e^{-2m\pi i x/\rho_k} \qquad (k = 1, 2, \cdots, n;\ m = 1, 2, 3, \cdots) \tag{3.13}$$

let us associate its reciprocal and let us call this associated function the adjoint of the given function. If we multiply any whatever function of the set (3.13) by the adjoint of any other function in the set, we have a product of the form

$$\prod_{k=1}^{n} e^{2l_k \pi i x/\rho_k}$$

where at least one and not more than two of the integers l_k are different from zero. There is a side of the polygon P on which x/ρ_k ranges from 0 to 1; on that side we denote x/ρ_k by λ_k. Then

$$\int_0^1 \int_0^1 \cdots \int_0^1 \left(\prod_{k=1}^{n} e^{2l_k \pi i \lambda_k} \right) d\lambda_1 d\lambda_2 \cdots d\lambda_n = 0. \tag{3.14}$$

If a like integral is formed with a function of the set (3.13) and the adjoint of that function then the integral corresponding to (3.14) has the value 1. Hence we have conditions of biorthogonality generalizing those pertaining to the case of Fourier series, here arising when $n=1$. Consequently we have a formal method of determining the coefficients in series (3.12) for a much more extensive class of functions than those for which we have already established the validity of such expansions. This suggests the generalization of the whole

* Series similar to those in (3.12) have been treated by P. Bohl, Magisterdissertation, Dorpat, 1893, and Journal für Mathematik, vol. 131 (1906), pp. 286–321.

theory of Fourier series to the particular class of series in (3.12) if not indeed to the more general class of §3.3; but we shall not now pursue these generalizations.

Other generalizations of the whole theory developed in this part of the paper will readily occur to the reader, including among others such extensions of the Birkhoff expansion theory as are parallel to the foregoing extension of the theory of Fourier series and also the extensions of these theories to the expansions of functions of several variables; but these also we leave to a future investigation.

3.5. **Applications to Bernoulli polynomials.** The theory in §3.4 affords elegant expansions of Bernoulli polynomials of higher order, namely, the polynomials B defined by the identity

$$\frac{\rho_1\rho_2 \cdots \rho_n t^n e^{xt}}{(e^{\rho_1 t}-1)\cdots(e^{\rho_n t}-1)} = \sum_{\nu=0}^{\infty} \frac{t^\nu}{\nu!} B_\nu^{(n)}(x \mid \rho_1, \cdots, \rho_n). \tag{3.15}$$

From this identity we have

$$B_\nu^{(n)}(x \mid \rho_1, \cdots, \rho_n) = \frac{\nu!\rho_1\rho_2\cdots\rho_n}{2\pi i}\int_C \frac{t^\nu e^{xt}}{(e^{\rho_1 t}-1)\cdots(e^{\rho_n t}-1)}\,\frac{dt}{t^{\nu+1}} \tag{3.16}$$

where C denotes a small circle about the point zero. Our theory is effective for values of ν not less than n.

Thus we have in particular the expansion

$$\begin{aligned} B_\nu^{(2)}(x \mid 1, i) = \frac{\nu!}{(2\pi)^{\nu-1}} \sum_{k=1}^{\infty} \frac{1}{k^{\nu-1}} \biggl\{ &\frac{e^{2k\pi x}}{e^{2k\pi}-1} + \frac{e^{-2k\pi x}}{e^{-2k\pi}-1} + (-i)^{\nu-2}\frac{e^{2k\pi i x}}{e^{-2k\pi}-1} \\ &+ (-i)^{\nu-2}\frac{e^{-2k\pi i x}}{e^{2k\pi}-1}\biggr\} \qquad (\nu = 3, 4, 5, \cdots). \end{aligned} \tag{3.17}$$

According to the general theory this series must converge for those values x, $x=u+iv$, for which u and v run independently over the closed interval (01). By considering separately the four cases $u<0$, $u>1$, $v<0$, $v>1$, it is easily shown that the series diverges in each case through having terms in the brackets become infinite in an exponential way as k becomes infinite. Hence the whole region of convergence of the series is the square whose vertices are $0, 1, 1+i, i$.

From this example it follows that the polygon P of convergence in the case of the general theory can not be extended to a larger region in which the series always converges.

UNIVERSITY OF ILLINOIS,
URBANA, ILL.

CONVERGENCE CRITERIA FOR DOUBLE FOURIER SERIES*

BY

J. J. GERGEN†

1.1. **Introduction.** We shall consider here analogues for double Fourier series of certain convergence criteria for simple Fourier series. The tests for simple series in question are the familiar tests of Dini, Jordan, de la Vallée-Poussin, Lebesgue, Young, and Hardy and Littlewood, and the tests obtained by various authors in generalizing the Young and the Lebesgue test. All these criteria are stated, the logical relations between them discussed, and references to them given in the author's paper **4**.‡ Rather than duplicate this material here we refer the reader to that paper. Analogues for some of these tests have appeared in the literature. Our first purpose here is to establish analogues of those remaining. Our second purpose is to discuss the logical relations between the tests for double series. We obtain, incidentally, an extension of Tonelli's convergence criterion for double series which deals with functions of bounded variation. Statements of our results and a general summary of the convergence theory are to found in §§1.2 to 1.6, the proofs of our theorems, in §§2.0 to 13.1. We do not always attempt to model the proof of a generalization after the proof of the original; but deduce first a test of the Lebesgue type, and from it the other tests. We thus obtain at the same time information as to the relations between the tests.

1.2. We suppose once and for all that the double Fourier series in question is that of an even-even function $f(u, v)$ which is integrable in the Lebesgue sense over the square $Q(0, 0; \pi, \pi)$ and is doubly periodic with period 2π in each variable. Further, we shall confine our attention to the behavior of the Fourier series of f at the origin. We have, then,

$$f(u, v) \sim \sum_{m,n=0}^{\infty} \lambda_{m,n} a_{m,n} \cos mu \cos nv,$$

where

* Presented to the Society, March 25, 1932; received by the editors May 25, 1932, and, in revised form, August 22, 1932.

† A number of the results contained in this paper were obtained while the author was a National Research Fellow. The problem of obtaining a generalization to double series of the Lebesgue test for simple series was suggested to the author by Professor Hardy; and the author wishes to thank him for this and other suggestions. The author also wishes to thank Dr. Agnew for reading the manuscript of this paper and suggesting several corrections and improvements.

‡ Numbers in bold face type refer to the Bibliography at the end of this paper.

$$\lambda_{0,0} = \tfrac{1}{4}, \quad \text{and } \lambda_{m,0} = \lambda_{0,n} = \tfrac{1}{2}, \quad \lambda_{m,n} = 1 \text{ for } 0 < m,\ 0 < n,$$

and

$$a_{m,n} = \frac{4}{\pi^2}\int\int_Q f(u, v) \cos mu \cos nv\, dudv.$$

The series for f at the origin is

$$\sum_{m,n=0}^{\infty} \lambda_{m,n}a_{m,n}; \tag{1.21}$$

the partial sum $s_{m,n}$ of order m, n of this series is

$$s_{m,n} = \sum_{i=0}^{m}\sum_{j=0}^{n}\lambda_{i,j}a_{i,j} = \frac{1}{\pi^2}\int\int_Q f(u, v)\frac{\sin (m+\frac{1}{2})u}{\sin \frac{1}{2}u}\ \frac{\sin (n+\frac{1}{2})v}{\sin \frac{1}{2}v}\, dudv;$$

and we are concerned with the limit

$$\lim_{m,n\to\infty} s_{m,n}$$

taken in the Pringsheim sense.* Any test for the convergence of the series (1.21) yields, of course, a test for convergence of the Fourier series of an arbitrary integrable function at an arbitrary point.

1.3. To simplify the writing we employ a form of the Landau limit notation. Given two functions $h(x, y)$ and $\psi(x, y)$, defined for all sufficiently small positive values of x and y, we write

$$h(x, y) = o\{\psi(x, y)\} \tag{1.31}$$

if, corresponding to each number $0<\epsilon$, we can choose $0<\delta_\epsilon$ so that

$$|h| \leqq \epsilon|\psi| \tag{1.32}$$

for $0<x\leqq\delta_\epsilon$, $0<y\leqq\delta_\epsilon$. We write

$$h(x, y) = O\{\psi(x, y)\} \tag{1.33}$$

if (1.32) holds for some ϵ and all sufficiently small positive values of x and y. Given two functions $h(x, y; k)$ and $\psi(x, y; k)$, defined for each large value of k for sufficiently small positive values of x and y, we write

$$h(x, y; k) = \bar{o}\{\psi(x, y; k)\} \tag{1.34}$$

* Pringsheim, 12, p. 103. The series (1.21) converges, to sum s, or

$$\lim_{m,n\to\infty} s_{m,n} = s$$

in the Pringsheim sense, if there corresponds to every number $0<\epsilon$ an integer N such that, if $N\leqq m$, $N\leqq n$, then $|s_{m,n}-s|\leqq\epsilon$.

if, corresponding to each $0<\epsilon$, we can choose, first, $0<k_\epsilon$; and then, $0<\delta_{k,\epsilon}$, so that (1.32) holds for

$$(1.35) \qquad 0 < x \leqq \delta_{k,\epsilon},\ 0 < y \leqq \delta_{k,\epsilon},\ k_\epsilon \leqq k.$$

We write

$$(1.36) \qquad h(x, y; k) = \bar{O}\{\psi(x, y; k)\}$$

if, corresponding to some ϵ, we can choose k_ϵ and $\delta_{k,\epsilon}$ as above so that (1.32) holds for all x, y, and k satisfying (1.35).†

1.4. The known tests for the convergence of the series (1.21) which are of interest here may now be recalled. In stating these, and in what follows, we understand that letters capped by bars, $(\bar{D})$, $(\bar{J})$, etc., have the same meanings as the same letters without the bars in the author's paper 4. Letters without bars refer to conditions and tests for double series. In some of the tests there are two or three conditions. We shall always regard the set of conditions in any test as a single condition and denote it by the same notation as we use to denote the test itself. Similarly, when a test involves but one condition we denote it in the same way as we do the test itself.

The conditions sufficient for the convergence of the series (1.21) *are in*

(D_Y) *Young's analogue of Dini's test* $(\bar{D})$:‡

$$\int_0^\pi \frac{du}{u}\int_0^\pi |\, f(u, v) - s - \xi_1(u) - \xi_2(v)\,| \frac{dv}{v} < \infty$$

where s is a constant and ξ_1, ξ_2 are functions such that $\xi_1(u)/u$, $\xi_2(v)/v$ are integrable over $(0, \pi)$;

---

† In case ψ is a non-vanishing function, then (1.31), (1.33), (1.34), (1.36) respectively holds if, and only if,

$$\operatorname*{limit}_{(x,y)\to(+0,+0)} (h/\psi) = 0, \qquad \operatorname*{\overline{limit}}_{(x,y)\to(+0,+0)} |h/\psi| < \infty,$$

$$\lim_{k\to\infty} \operatorname*{\overline{limit}}_{(x,y)\to(+0,+0)} |h/\psi| = 0, \quad \overline{\lim_{k\to\infty}} \operatorname*{\overline{limit}}_{(x,y)\to(+0,+0)} |h/\psi| < \infty.$$

‡ Young, 15, p. 182. About the same time as Young's paper appeared Küstermann, 10, p. 28, published an analogue of Dini's test. Later a test of this type was given by Merriman, 3, p. 129. The test (D_Y) is not exactly the test of any of these authors. It includes them all as particular cases however.

Young's condition is

(D_Y^*) *the function* $(f-s)/(uv)$ *is integrable over* Q.

Young proves that, if (D_Y^*) is satisfied, then so also is (V_Y^*) (see the footnote on (V_Y) above). Using Young's method it is not difficult to show that (D_Y) implies (V_Y^*). Thus (D_Y) is a sufficient condition for convergence.

(J_H) *Hardy's analogue of Jordan's test* $(\bar{J})$:†

(J_H') *f is finitely defined everywhere in Q and*

$$\int_0^\pi \int_0^\pi |d_{u,v}f(u, v)| < \infty,$$

where the integral represents the total variation of f over Q,‡ *and*

$$(J_H'') \qquad \int_0^\pi |d_u f(u, 0)| < \infty, \quad \int_0^\pi |d_v f(0, v)| < \infty,$$

where the first integral is the total variation of $f(u, 0)$, *and the second, the total variation of* $f(0, v)$, *over* $(0, \pi)$;§

(J_T) *Tonelli's analogue of* $(\bar{J})$:‖

(J_T') *f is finitely defined everywhere in Q and*

$$V_1(v) \equiv \int_0^\pi |d_u f(u, v)| < V(v), \quad V_2(u) \equiv \int_0^\pi |d_v f(u, v)| < V(u),$$

where V is integrable over $(0, \pi)$, *the first for every V, and the second for every u, on* $(0, \pi)$,¶ *and*

$$(J_T'') \qquad \begin{aligned} W_1(x, y) &\equiv \lim_{\tau\to+0} \int_\tau^x |d_u f(u, y)| = o(1), \\ W_2(x, y) &\equiv \lim_{\tau\to+0} \int_\tau^y |d_v f(x, v)| = o(1).^{**} \end{aligned}$$

---

† Hardy, **5**, p. 65.

‡ The total variation of f over the rectangle $(a_1, b_1; a_2, b_2)$,

$$\int_{a_1}^{a_2}\int_{b_1}^{b_2} |d_{u,v}f(u, v)|,$$

is defined as the upper bound of all sums of the type

$$\sum_{i=1}^{m} \sum_{j=1}^{n} |f(u_i, v_j) - f(u_i, v_{j-1}) - f(u_{i-1}, v_j) + f(u_{i-1}, v_{j-1})|$$

where u_i, $i=0, 1, \cdots, m$, and v_j, $j=0, 1, \cdots, n$, are any numbers such that $a_1 = u_0 < u_1 < \cdots < u_m = a_2$, $b_1 = v_0 < v_1 < \cdots < v_n = b_2$.

§ Hardy states this second condition as

$(J_H''^*)$ $\int_0^\pi |d_u f(u, v)| < \infty$, $\int_0^\pi |d_v f(u, v)| < \infty$,

the first for every v, the second for every u, on $(0, \pi)$.

It is pointed out by Young, **15**, p. 142, that, if (J_H') and (J_H'') hold, then $(J_H''^*)$ holds.

‖ Tonelli, **13**, p. 455, and **14**.

¶ This condition is stated by Tonelli as

$(J_T'^*)$ V_1, V_2 *are integrable over* $(0, \pi)$.

Since, as Professor Adams pointed out to the author, there are functions for which V_1 and V_2 are not measurable, there is some gain in generality in taking the condition as we do. That (J_T') and (J_T'') are sufficient conditions for convergence, we prove by Theorems I, II and III below.

When a function satisfies (J_H) it may be said to be of bounded variation in the Hardy sense, and when it satisfies (J_T'), of bounded variation in the Tonelli sense. Other definitions of bounded variation have been given by various authors. For a complete discussion of these, see Adams, **1**. For a convergence theorem involving another definition, see Hobson, **2**, p. 705. For further references on convergence criteria, see Tonelli, **14**.

** Tonelli states this condition in a slightly different but equivalent way.

(V_Y) *Young's analogue of de la Vallée-Poussin's test* ($\overline{V}$):† *the mean value*

$$F(u, v) = \frac{1}{uv}\int_0^u d\sigma \int_0^v f(\sigma, t)dt \tag{1.41}$$

of f satisfies (J_H).

In (D_Y) *the sum of the series is s, in* (J_H) *and* (J_T), $f(+0, +0)$, *and in* (V_Y), $F(+0, +0)$.

Each of the above tests is plainly analogous to the corresponding test for simple series. There is, however, one aspect in which these tests and, in fact, all the tests given here for double series, differ from the original tests. In each test for double series there is some condition on f, other than integrability, over the whole square Q, whereas for simple series, the only conditions imposed, other than integrability, are neighborhood conditions. Now, by the analogue of the Riemann-Lebesgue theorem,‡ the behavior of f in the square $(\delta, \delta; \pi, \pi)$, provided $0<\delta$, has no effect on the convergence of the series (1.21). Thus this difference could be partially eliminated; but we cannot, as might be expected, confine our conditions to neighborhood conditions. Conditions on f in the "cross-neighborhood" of the origin are essential. Some of the above tests were originally stated with only cross-neighborhood conditions, and we could state those which follow thus. We state the tests as we do for simplicity.

From each of the above tests the corresponding test for simple series can readily be deduced. We have, when f is a function of u alone, $f=\bar{f}(u)$, say,

$$s_{m,n} = s_{m,0} = \frac{1}{\pi}\int_0^\pi \bar{f}(u)\frac{\sin (m+\frac{1}{2})u}{\sin \frac{1}{2}u}du,$$

which is the mth partial sum of the simple Fourier series of $\bar{f}$ corresponding to the point $u=0$. Now, if $\bar{f}$ satisfies ($\overline{D}$), f satisfies (D_Y); and if $\bar{f}$ satisfies ($\overline{V}$), f satisfies (V_Y). Hence from (D_Y) we can immediately deduce ($\overline{D}$), and from (V_Y), ($\overline{V}$). The passage from (J_H) and (J_T) to ($\overline{J}$) is not so immediate, but a simple application of the Riemann-Lebesgue theorem leads directly to the desired conclusion.

1.5. An examination of §1.4 reveals that the types of tests for simple series which have not been considered for double series are Young's, Hardy

† Young, 15, p. 170. Young states his condition in another but equivalent form, namely: (V_Y^*) *Fuv* csc *u* csc *v satisfies* (J_H).

The right-hand member of (1.41) has, of course, no meaning when $u=0$ or when $v=0$. It is implied that F can be so defined on the axes as to satisfy (J_H).

‡ For this analogue, see Young, 15, p. 138.

and Littlewood's, and Lebesgue's. Listed below is our extension of Tonelli's criterion and the analogues we obtain of tests of these types.

The conditions sufficient for the convergence of the series (1.21), *to sum s, are, in*

(J_R) *our extension of Tonelli's test* (J_T): (J_T'),

$$(J_R'') \qquad W_1(x, y) = O(1), \; W_2(x, y) = O(1),$$

and

$$(C_1) \qquad \phi_1(x, y) \equiv \int_0^x du \int_0^y \phi(u, v)dv = o(xy),$$

where $\phi = f - s$;

(Y) *our analogue of Young's test* $(\overline{Y})$:

(Y') *f is finitely defined everywhere in Q and*

$$(1.51) \qquad \int_0^x \int_0^y | d_{u,v}\{uvf(u, v)\} | < Axy \text{ for } 0 < x \leqq \pi, 0 < y \leqq \pi,$$

where A is independent of x and y, and

$$(C_0) \qquad \phi(x, y) = o(1);$$

(Y_P) *our analogue of Pollard's generalization* $(\overline{Y}_{\overline{P}})$ *of* $(\overline{Y})$: (Y') *and* (C_1);

(HL) *our analogue of Hardy and Littlewood's test* $(\overline{HL})$:

$$(HL') \qquad \int_0^{\pi-x} du \int_0^{\pi-y} | \Delta_{x,y}f(u, v) |^{p_1}dv = O(xy),$$

where

$$\Delta_{x,y} f(u, v) = f(u+x, v+y) - f(u+x, v) - f(u, v+y) + f(u, v),$$

for some $1 \leqq p_1$,

$$(HL'') \qquad \int_0^x du \int_0^{\pi-y} | \Delta_y f |^{p_2} dv = O(xy), \quad \int_0^y dv \int_0^{\pi-x} | \Delta_x f |^{p_3} du = O(xy),$$

where†

$$\Delta_y f = f(u, v + y) - f(u, v), \quad \Delta_x f = f(u + x, v) - f(u, v),$$

for some $1 \leqq p_2$ *and some* $1 \leqq p_3$, *and* (C_1);

† There is some confusion in the notation $\Delta_x f$, $\Delta_y f$, and $\Delta_{x,y} f$, but this is not serious. Whenever we have $\Delta_x h$, h has u as one of its arguments, and $\Delta_x h$ is the first difference

$$\Delta_x h = h(u+x \cdots) - h(u \cdots).$$

Similarly, y always appears with and is coupled with v. Finally, whenever we have $\Delta_{x,y} h$, h has u and v as two of its arguments and $\Delta_{x,y} h$ is the first double difference

$$\Delta_{x,y} h = h(u+x \cdots v+y \cdots) - h(u+x \cdots v \cdots) - h(u \cdots v+y \cdots) + h(u \cdots v \cdots).$$

(L_1) *our analogue of Lebesgue's test* $(\bar{L}_1)$:

$$(L_1') \qquad \int_x^{\pi-x} \frac{du}{u} \int_y^{\pi-y} |\Delta_{x,y} f| \frac{dv}{v} = o(1),$$

$$(L_1'') \qquad \int_0^x du \int_y^{\pi-y} |\Delta_y f| \frac{dv}{v} = o(x), \quad \int_0^y dv \int_x^{\pi-x} |\Delta_x f| \frac{du}{u} = o(y),$$

and

$$(C_1^*) \qquad \phi_1^*(x, y) \equiv \int_0^x du \int_0^y |\phi(u, v)| \, dv = o(xy);$$

(L_2) *our analogue of Lesbegue's test* $(\bar{L}_2)$:

$$(L_2') \qquad \zeta_1(x, y) \equiv \int_x^{\pi-x} du \int_y^{\pi-y} \left| \Delta_{x,y} \left\{ \frac{\phi(u, v)}{uv} \right\} \right| dv = o(1),$$

and

$$(L_2'') \qquad \begin{aligned} \xi_1 &\equiv \int_0^x du \int_y^{\pi-y} \left| \Delta_y \left\{ \frac{\phi}{v} \right\} \right| dv = o(x), \\ \eta_1 &\equiv \int_0^y dv \int_x^{\pi-x} \left| \Delta_x \left\{ \frac{\phi}{u} \right\} \right| du = o(y); \end{aligned}$$

(L_P) *our analogue of Pollard's generalization* $(\bar{L}_P)$ *of* $(\bar{L}_2)$:

$$(L_p') \qquad \zeta(x, y; k) \equiv \int_{kx}^{\pi-x} du \int_{ky}^{\pi-y} \left| \Delta_{x,y} \left\{ \frac{\phi}{uv} \right\} \right| dv = \bar{o}(1),$$

$$(L_p'') \qquad \begin{aligned} \xi &\equiv \int_0^x du \int_{ky}^{\pi-y} \left| \Delta_y \left\{ \frac{\phi}{v} \right\} \right| dv = \bar{o}(x), \\ \eta &\equiv \int_0^y dv \int_{kx}^{\pi-x} \left| \Delta_x \left\{ \frac{\phi}{u} \right\} \right| du = \bar{o}(y), \end{aligned}$$

and (C_1);

(L_R) *our analogue of Gergen's generalization* $(\bar{L}_R)$ *of* $(\bar{L}_1)$:

$$(L_R') \qquad \gamma(x, y; k) \equiv \int_{kx}^{\pi-x} \frac{du}{u} \int_{ky}^{\pi-y} |\Delta_{x,y} f| \frac{dv}{v} = \bar{o}(1),$$

$$(L_R'') \qquad \begin{aligned} \alpha &\equiv \int_0^x du \int_{ky}^{\pi-y} |\Delta_y f| \frac{dv}{v} = \bar{o}(x), \\ \beta &\equiv \int_0^y dv \int_{kx}^{\pi-x} |\Delta_x f| \frac{du}{u} = \bar{o}(y), \end{aligned}$$

and (C_1).

From each of these tests the corresponding test for simple series can immediately be deduced; but it will be noticed that the most general continuity condition we use is (C_1) and not the analogue of $(\overline{C})$, namely:
(C) *the series* (1.21) *is summable, to sum s, by some Cesàro means.*
Thus we fail to extend completely to double series the tests $(\overline{HL})$ and $(\overline{L}_R)$, and we fail to obtain any analogue of Hardy and Littlewood's generalization $(\overline{Y}_{HL})$ of $(\overline{Y})$ other than (Y_P). The problem remains unsolved whether we can replace (C_1) by (C) in (L_R), (Y_P), and (HL). This problem, if one follows the ideas in simple series, involves proving that the characteristic† conditions of these tests imply the equivalence of (C_1) and (C), and this in turn involves obtaining a generalization to double series of Hardy and Littlewood's‡ theorem on summability and continuity. Of course the fact that each of the above tests leads directly to the corresponding test for simple series is due largely to the equivalence of $(\overline{C}_1)$ to $(\overline{C})$ whenever the latter is used.

To establish the above tests we deduce first (L_R) and then show that (L_R) contains all the other tests as particular cases. The facts in regard to (L_R) we state for convenience in the following theorem, the proof of which is to be found in §§2.0 to 3.1. The facts in regard to the relation of (L_R) to the rest of the tests are given in Theorems II to VII below.

THEOREM I. *If* (L_R) *holds, then the series* (1.21) *converges, to sum s.*

1.6. Turning now to the logical relations§ between the tests, we first state the following theorems, the proofs of which are to be found in §§4.1 to 13.1.

THEOREM II. (a) *The conditions* (L_1) *and* (L_2) *are equivalent.* (b) *The condition* (L_2) *implies* (L_P), *while* (L_P) *implies* (L_2) *if* (C_1^*) *holds.* (c) *The condition* (L_P) *implies* (L_R), *while* (L_R) *implies* (L_P) *if*

$$\phi_1^*(x, y) = O(xy). \tag{1.61}$$

Thus (L_1), (L_2), (L_P), *and* (L_R) *are equivalent if* (C_1^*) *holds.*‖

† The characteristic condition of a test consists of the conditions individual to the test. It is to be distinguished from the continuity condition, which is either (C_0) or a generalization of (C_0). In (L_R), for example, the characteristic condition is $(L_R')+(L_R'')$.

‡ Hardy and Littlewood, 7.

§ All our conclusions here are to the effect that certain conditions imply others. We make no attempt to prove that the implications stated are not reversible. For some examples of this type, see Hardy, 6. Hardy's examples are for simple series, but conclusions for double series can easily be deduced from them.

In discussing these relations it is well to point out again that, because one test is included in another, the latter is not for that reason a better test. If one reasoned in that way the best test would be the one in which the condition for convergence is that the series converge.

‖ This theorem contains some new information for simple series: that $(\overline{L}_P)$ implies $(\overline{L}_2)$ if (C_1^*) holds, and that $(\overline{L}_R)$ implies $(\overline{L}_P)$ if the condition corresponding to (1.61) holds.

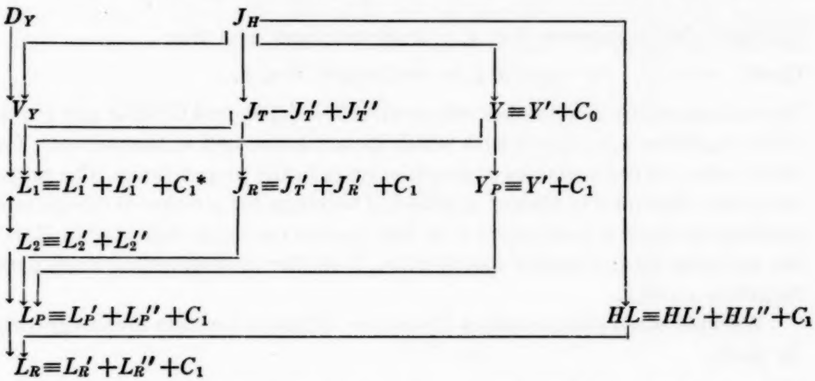
THEOREM III. *The condition* (J_T) *implies* (J_R) *and* (L_1), *both with* $s=f(+0, +0)$. *Moreover*, (J_R) *implies* (L_P).

THEOREM IV. *The condition* (Y) *implies* (Y_P) *and* (L_1), *while* (Y_P) *implies* (L_P).

THEOREM V. *The condition* (HL) *implies* (L_R).

THEOREM VI. *The condition* (J_H) *implies* (Y) *and* (HL), *both with* $s=f(+0, +0)$, *and the latter with* $p_1=p_2=p_3=1$.

THEOREM VII. *The condition* (V_Y) *implies* (L_1) *with* $s=F(+0, +0)$.



Combining these results with the fact that both (J_H) and (D_Y) imply (V_Y),† and (J_H) implies (J_T),‡ we obtain the accompanying diagram. In this diagram a directed line running from one letter to another indicates that the condition represented by the former implies that represented by the latter. In any implication in which s occurs only in the implied condition, s is understood to have the value indicated in the above theorem concerning this implication. It should be noted that, aside from the differences due to our use of (C_1) rather than (C), there are only two essential differences between this diagram and the author's§ diagram for simple series: first, we do not indicate here any implications between characteristic conditions; and secondly, we have here two new conditions, (J_T) and (J_R). In regard to the first difference it might be pointed out that our proofs show that all implications indicated for simple series carry over to double series, and thus that, in particular, the characteristic condition of (L_R) is implied by every other

† See the footnote on (D_Y) and Young, 15, p. 181.

‡ Tonelli, 13, p. 470.

§ Gergen, 4, p. 257.

characteristic condition. In regard to (J_T) and (J_R), these conditions, while analogous to $(\bar{J})$ in some respects, do not seem to be contained in (V_Y), (Y), and (HL). For this reason, and also because of their general character, these conditions seem to be essentially connected with a space of higher dimensionality than one.

2.0. **Lemmas for Theorem I.** The proof of Theorem I rests on the following lemmas. In these lemmas and throughout the rest of the paper we write for convenience

$$K(u, v; x, y) = K(u, v) = \sin\frac{\pi u}{x}\csc\frac{1}{2}u\sin\frac{\pi v}{y}\csc\frac{1}{2}v.$$

We shall always suppose that x, y, k are numbers such that

$$0 < x \leqq \pi,\ 0 < y \leqq \pi,\ 0 < k. \tag{2.01}$$

We understand by A a number whose value is independent of all or any group of the variables u, v, x, y, k with which we are concerned at the moment, for those values of the variables in question lying in the proper range. The range for u and that for v is always specified. The range for x either is completely specified or else it is understood to be that part of the range indicated in (2.01) not excluded by any partial specification. A similar understanding holds with regard to y and k.

We shall often have occasion to use the following formula for integrating by parts:

$$\begin{aligned}\int_{a_1}^{a_2} du \int_{b_1}^{b_2} \rho\psi'(u)\psi''(v)dv &= \rho_1(a_2, b_2)\psi(a_2,b_2) - \int_{a_1}^{a_2} \rho_1(u, b_2)\psi_u(u, b_2)du \\ &\quad - \int_{b_1}^{b_2} \rho_1(a_2, v)\psi_v(a_2, v)dv + \int_{a_1}^{a_2} du \int_{b_1}^{b_2} \rho_1\psi_{uv}dv,\end{aligned} \tag{2.02}$$

where

$$\psi(u, v) = \psi'(u)\psi''(v),$$

$$\rho_1(u, v) = \int_{a_1}^{u} d\sigma \int_{b_1}^{v} \rho(\sigma, t)dt.$$

This formula is valid if ρ is integrable on $(a_1, b_1; a_2, b_2)$, ψ' is absolutely continuous on (a_1, a_2), and ψ'' is absolutely continuous on (b_1, b_2).†

† This formula can readily be established by a double application of the formula for integrating by parts an integral involving but one variable and the application of other familiar results in the theory of Lebesgue integrals. The only question likely to occur is that of the measurability of the function $\int_{b_1}^{v}\rho(u, t)dt$ and this question is answered in a theorem of Carathéodory, **2**, p. 656. In any case the formula is a particular case of one given by Hobson, **8**, p. 666.

We have

$$s_{m,n} - s = \frac{1}{\pi^2}\int_0^\pi du \int_0^\pi \phi(u, v)K\{u, v; \pi/(m + \tfrac{1}{2}), \pi/(n + \tfrac{1}{2})\}dv,$$

and our problem is to show that, if (L_R) holds, then

$$S(x, y) \equiv \int_0^\pi du \int_0^\pi \phi K(u, v; x, y) = o(1).$$

As in the proof of $(\overline{L}_R)$ the problem is solved by breaking this integral into several parts and considering each part separately. In the lemmas we consider integrals over the region "near" the boundary of Q and also several functions which occur in the treatment of the integral over the area "away" from the boundary.

2.1. Lemma 1. *If* (C_1) *holds and if* $0<a$, $0<b$, *then*

$$I_1(x, y) \equiv \int_0^{ax} du \int_0^{by} \phi K dv = o(1). \tag{2.11}$$

Further,

$$I_2 \equiv \int_{\pi-ax}^{\pi} du \int_{\pi-by}^{\pi} \phi K dv = o(1)$$

under the sole assumption that ϕ *is integrable in* Q.

We have

$$\left.\begin{array}{ll} uv\,|\,K\,| < A, & uvx\,|\,K_u\,| < A \\ uvy\,|\,K_v\,| < A, & uvxy\,|\,K_{uv}\,| < A \end{array}\right\} \text{ for } 0 < u \leqq \pi,\ 0 < v \leqq \pi. \tag{2.12}$$

Further, applying (2.02),

$$I_1 = \phi_1(ax, by)K(ax, by) - \int_0^{ax} \phi_1(u, by)K_u(u, by)du$$
$$- \int_0^{by} \phi_1(ax, v)K_v(ax, v)dv + \int_0^{ax} du \int_0^{by} \phi_1 K_{uv} dv.$$

Thus

$$I_1 = O\left\{ \underset{0<u\leqq ax, 0<v\leqq by}{\text{maximum}} \left| \phi_1(u, v)/(uv) \right| \right\} = o(1)$$

by (C_1). This is (2.11).

As for I_2, we have immediately, since ϕ is integrable in Q,

$$I_2 = O\left\{ \int_{\pi - ax}^{\pi} du \int_{\pi - by}^{\pi} |\phi| \, dv \right\} = o(1).$$

2.2. Lemma 2. *If* (C_1) *and* (L_R'') *hold, then*

$$\int_0^x du \int_{\pi - y}^{\pi} \phi dv = o(x), \tag{2.21}$$

$$|\phi_1(x, y)| < Axy. \tag{2.22}$$

We observe first that, corresponding to an arbitrary number $0 < \epsilon$, we can choose $0 < k_0$ and $0 < \delta < \pi/2$ so that

$$\begin{gathered} |\phi_1(x, y)| < \epsilon xy \text{ for } x \leqq \delta, \ y \leqq \delta, \\ \alpha(x, y; k_0) < \epsilon x, \quad \beta(x, y; k_0) < \epsilon y \text{ for } x \leqq \delta, \ y \leqq \delta. \end{gathered} \tag{2.23}$$

We observe next that, if x, c, and z are numbers such that

$$x \leqq \delta, \ (k_0 + 1)z \leqq \delta \leqq c < c + z \leqq \pi, \tag{2.24}$$

we consequently have

$$\begin{aligned} \left| \int_0^x du \int_c^{c+z} \phi dv \right| &= \left| \int_0^x du \left\{ \int_{(k_0+1)z}^{c+z} - \int_{k_0 z}^{c} + \int_{k_0 z}^{(k_0+1)z} \right\} \phi dv \right| \\ &\leqq \int_0^x du \int_{k_0 z}^{\pi - z} |\phi(u, v + z) - \phi(u, v)| \, dv + |\phi_1\{x, (k_0 + 1)z\}| \\ &\quad + |\phi_1(x, k_0 z)| \\ &< \pi\epsilon x + \epsilon x(k_0 + 1)z + \epsilon x k_0 z < 2\pi\epsilon x. \end{aligned} \tag{2.25}$$

We can now easily prove that (2.21) holds. In fact, if $x \leqq \delta$ and $(k_0+1)y \leqq \delta$, then (2.24) holds with $c = \pi - y$ and $z = y$; and hence, applying (2.25),

$$\left| \int_0^x du \int_{\pi - y}^{\pi} \phi \, dv \right| < 2\pi\epsilon x.$$

Since ϵ was arbitrary, this proves that (2.21) holds.

As for (2.22), let us first suppose that $x \leqq \delta < y$. Then, choosing z so that $N = (\pi - \delta)/z$ is an integer and so that $0 < (k_0+1)z \leqq \delta$, and denoting by a the largest of the numbers $\delta, \delta + z, \cdots, \pi - z$ less than y, we have

$$\begin{aligned} |\phi_1(x, y)| &\leqq |\phi_1(x, \delta)| + \sum_{n=0}^{N-1} \left| \int_0^x du \int_{\delta + nz}^{\delta + (n+1)z} \phi \, dv \right| + \left| \int_0^x du \int_a^y \phi \, dv \right| \\ &< \epsilon x\delta + 2\pi N\epsilon x + 2\pi\epsilon x \end{aligned}$$

upon applying (2.23) and (2.25). Thus, for $x \leqq \delta < y$, we have

$$|\phi_1(x, y)| < Axy. \tag{2.26}$$

Similarly, (2.26) holds for $y \leqq \delta < x$. Accordingly, because of (2.23) and the obvious fact that (2.26) holds for $\delta \leqq x$, $\delta \leqq y$, the proof is complete.

2.3. LEMMA 3. *If* (C_1) *and* (L_R'') *hold and if* $0 < a$, $0 < b$, *then*

$$I_3 \equiv \int_0^{ax} du \int_{\pi - by}^{\pi} \phi K \, dv = o(1). \tag{2.31}$$

Writing

$$\bar{\phi}(u, v) = \int_0^u d\sigma \int_{\pi - by}^{v} \phi(\sigma, t) dt,$$

we have

$$I_3 = \bar{\phi}(ax, \pi) K(ax, \pi) - \int_0^{ax} \bar{\phi}(u, \pi) K_u(u, \pi) du$$

$$- \int_{\pi - by}^{\pi} \bar{\phi}(ax, v) K_v(ax, v) dv + \int_0^{ax} du \int_{\pi - by}^{\pi} \bar{\phi} K_{uv} dv$$

$$= O(\text{maximum} \, |\bar{\phi}(u, v)/v|) \qquad (0 < u \leqq ax, \pi - by \leqq v \leqq \pi).$$

Using Lemma 2 now we conclude the proof.

2.4. LEMMA 4. *If* (C_1) *and* (L_R'') *hold and if* $0 < a$, *then*

$$J_1(x, y; k) \equiv \int_0^{ax} du \int_{ky}^{\pi - 2y} \phi \Omega \, dv = \bar{o}(1),$$

where

$$\Omega = \Omega(u, v; x, y) = \Omega(u, v) = \sin \frac{\pi u}{x} \csc \frac{1}{2} u \sin \frac{\pi v}{y} \, \omega(v; y),$$

and

$$\omega(v; y) = 2 \csc \tfrac{1}{2}(v + y) - \csc \tfrac{1}{2}(v + 2y) - \csc \tfrac{1}{2} v.$$

Writing

$$\bar{\phi}(u, v) = \int_0^u d\sigma \int_{ky}^{v} \phi(\sigma, t) dt,$$

we have, by Lemma 2,

$$|\bar{\phi}(u, v)| < Auv \quad \text{for } 0 < u \leqq \pi, \ ky \leqq v \leqq \pi.$$

On the other hand, we have

$$(2.41)\qquad v^3\,|\,\omega(v;\,y)\,|\,<Ay^2,\qquad v^4\,|\,\omega_v\,|\,<Ay^2 \text{ for } y<v\leqq\pi-y;\dagger$$

and thus

$$(2.42)\qquad \left.\begin{aligned} uv^3\,|\,\Omega(u,v)\,|\,<Ay^2,\ uv^3x\,|\,\Omega_u\,|\,<Ay^2\\ uv^3\,|\,\Omega_v\,|\,<Ay,\ uv^3x\,|\,\Omega_{uv}\,|\,<Ay\end{aligned}\right\}$$

$$\text{for } 0<u\leqq\pi,\ y\leqq v\leqq\pi-y.$$

Thus

$$\begin{aligned} J_1 &= \bar{\phi}(ax,\pi-2y)\Omega(ax,\pi-2y)-\int_0^{ax}\bar{\phi}(u,\pi-2y)\Omega_u(u,\pi-2y)du\\ &\quad-\int_{ky}^{\pi-2y}\bar{\phi}(ax,v)\Omega_v(ax,v)dv+\int_0^{ax}du\int_{ky}^{\pi-2y}\bar{\phi}\Omega_{uv}dv\\ &= \bar{O}(y^2+y^2+1/k+1/k)=\bar{o}(1);\end{aligned}$$

which proves the lemma.

2.5. Lemma 5. *If* (C_1) *and* (L_R'') *hold and if* $0<a$, *then*

$$(2.51)\qquad I_4\equiv\int_0^{ax}du\int_0^{\pi}\phi K\,dv=o(1).$$

We have, by Lemmas 1, 3, and 4,

$$\begin{aligned} 4I_4 &= \int_0^{ax}du\left\{\int_{ky}^{\pi-2y}+2\int_{(k+1)y}^{\pi-y}+\int_{(k+2)y}^{\pi}\right\}\phi K\,dv+\bar{o}(1)\\ &= -J_1+J_1'+\bar{o}(1)=J_1'+\bar{o}(1),\end{aligned}$$

where

$$J_1'=\int_0^{ax}\frac{\sin\dfrac{\pi u}{x}}{\sin\frac{1}{2}u}du\int_{ky}^{\pi-2y}\left\{\frac{\Delta_{2y}\phi}{\sin\frac{1}{2}(v+2y)}-\frac{2\Delta_y\phi}{\sin\frac{1}{2}(v+y)}\right\}\sin\frac{\pi v}{y}\,dv.$$

But

$$\begin{aligned} J_1' &= \bar{O}\left\{\frac{1}{x}\int_0^{ax}du\int_{ky}^{\pi-2y}\left|\frac{\Delta_{2y}\phi}{\sin\frac{1}{2}(v+2y)}-\frac{2\Delta_y\phi}{\sin\frac{1}{2}(v+y)}\right|dv\right\}\\ &= \bar{O}\left\{\alpha(ax,2y;\tfrac{1}{2}k)/x+\alpha(ax,y;k)/x\right\}=\bar{o}(1).\end{aligned}$$

Since I_4 is independent of k, the lemma follows.

† Gergen, 4, p. 271. The first inequality in (2.41) is (8.72) and the second is (8.73) of that paper with $M=A$, $m=2$, $\nu=1$, $\rho=0$, $t=v$.

2.6. LEMMA 6. *If* (L_R) *and* (L_R'') *hold, then*

$$J_2 \equiv \int_{\pi-y}^{\pi} dv \int_{kx}^{\pi-x} |\Delta_x\phi| \frac{du}{u} = \bar{o}(1) \tag{2.61}$$

and $0<x_0<\pi$ *and* $0<k_0$ *can be found so that*

$$\beta(x, y; k) < Ay \qquad \text{for } x \leqq x_0,\ k_0 \leqq k. \tag{2.62}$$

We first observe that, corresponding to an arbitrary number $0<\epsilon$, we can choose $0<k_0$ and x_0 so that

$$0 < x_0(k_0+1) < \pi/2, \tag{2.63}$$

$$\beta(x, y; k_0) < \epsilon y, \qquad \gamma(x, y; k_0) < \epsilon \text{ for } x \leqq x_0,\ y \leqq x_0. \tag{2.64}$$

We next observe that, if

$$x \leqq x_0,\ (k_0+1)z \leqq x_0 \leqq c < c+z \leqq \pi, \tag{2.65}$$

we consequently have

$$\begin{aligned}\int_c^{c+z} dv \int_{k_0x}^{\pi-x} |\Delta_x\phi| \frac{du}{u} &= \left\{ \int_{(k_0+1)z}^{c+z} - \int_{k_0x}^{c} + \int_{k_0x}^{(k_0+1)z} \right\} \int_{k_0x}^{\pi-x} |\Delta_x\phi| \frac{du}{u} \\ &\leqq \pi\gamma(x, z; k_0) + \beta\{x, (k_0+1)z; k_0\} \\ &< \pi\epsilon + \epsilon(k_0+1)z < 2\pi\epsilon.\end{aligned} \tag{2.66}$$

Consider, then, (2.61). If $x\leqq x_0$ and $(k_0+1)y\leqq x_0$, then (2.65) holds with $c=\pi-y$ and $z=y$. Accordingly, because of (2.66) and the fact that

$$J_2(x, y; k) \leqq J_2(x, y; k_0) \text{ for } k_0 \leqq k,$$

we have

$$J_2 < 2\pi\epsilon \text{ for } x \leqq x_0,\ (k_0+1)y \leqq x_0,\ k_0 \leqq k.$$

Since ϵ was arbitrary, this proves that (2.61) holds.

As for (2.62), we observe that, because of (2.64) and the fact that

$$\beta(x, y; k) \leqq \beta(x, y; k_0) \text{ for } k_0 \leqq k,$$

it is enough to prove that

$$\beta(x, y; k_0) < Ay \text{ for } x \leqq x_0 < y.$$

But this is immediate; for, choosing z so that $N=(\pi-x_0)/z$ is an integer and so that $0<(k_0+1)z\leqq x_0$, we have

$$\beta(x, y; k_0) \leqq \beta(x, \pi; k_0) \leqq \beta(x, x_0; k_0) + \sum_{n=0}^{N-1} \int_{x_0+nx}^{x_0+(n+1)x} dv \int_{k_0 x}^{\pi - x} |\Delta_x \phi| \frac{du}{u}$$

$$< \epsilon x_0 + 2\pi N \epsilon < Ay$$

for $x \leqq x_0 < y$ by (2.64) and (2.66). This completes the proof.

2.7. LEMMA 7. *If* (L_R) *holds and if* $0 < b$, *then*

$$I_5 = \int_{\pi - by}^{\pi} dv \int_0^{\pi} \phi K \, du = o(1).$$

We have, using Lemmas 1 and 3,

$$4I_5 = \int_{\pi - by}^{\pi} dv \left\{ \int_{kx}^{\pi - 2x} + 2 \int_{(k+1)x}^{\pi - x} + \int_{(k+2)x}^{\pi} \right\} \phi K \, du + \bar{o}(1)$$

$$= J_3' + J_3'' + \bar{o}(1),$$

where

$$J_3' = \int_{\pi - by}^{\pi} \frac{\sin \dfrac{\pi v}{y}}{\sin \frac{1}{2} v} dv \int_{kx}^{\pi - 2x} \left\{ \frac{\Delta_{2x}\phi}{\sin \frac{1}{2}(u + 2x)} - \frac{2\Delta_x \phi}{\sin \frac{1}{2}(u + x)} \right\} \sin \frac{\pi u}{x} \, du,$$

$$J_3'' = - \int_{\pi - by}^{\pi} dv \int_{kx}^{\pi - x} \phi \Omega(v, u; y, x) du.$$

Now, as for J_3', we have immediately, upon applying Lemma 6,

$$J_3' = \bar{O}\{J_2(2x, by; \tfrac{1}{2}k) + J_2(x, by; k)\} = \bar{o}(1).$$

Finally, as for J_3'', upon setting

$$\bar{\phi}(u, v) = \int_{\pi - by}^{v} dt \int_{kx}^{u} \phi(\sigma, t) d\sigma,$$

we have

$$|\bar{\phi}(u, v)| < Au \text{ for } kx \leqq u \leqq \pi, 0 < \pi - by \leqq v \leqq \pi,$$

by Lemma 2. Thus, noting that (2.42) is applicable to the function

$$\bar{\Omega}(u, v) = \Omega(v, u; y, x),$$

we have

$$J_3'' = -\bar{\phi}(\pi - 2x, \pi)\bar{\Omega}(\pi - 2x, \pi) + \int_{kx}^{\pi-2x} \bar{\phi}(u, \pi)\bar{\Omega}_u(u, \pi)du$$

$$+ \int_{\pi-by}^{\pi} \bar{\phi}(\pi - 2x, v)\bar{\Omega}_v(\pi - 2x, v)dv - \int_{\pi-by}^{\pi} dv \int_{kx}^{\pi-2x} \bar{\phi}\bar{\Omega}_{uv}\, du$$

$$= \bar{O}(x^2 + 1/k + x^2 + 1/k) = \bar{o}(1).$$

This completes the proof.

2.8. Lemma 8. *If* (C_1) *and* (L_R'') *hold, then*

$$J_3 \equiv \int_{kx}^{\pi-2x} \omega(u; x) \sin \frac{\pi u}{x} du \int_{ky}^{\pi-2y} \omega(v; y)\phi(u + x, v + y) \sin \frac{\pi v}{y}\, dv = \bar{o}(1).$$

First, setting

$$w(u, v) = \omega(u; x) \sin (\pi u/x)\omega(v; y) \sin (\pi v/y),$$

and using (2.41), we have

$$\left.\begin{array}{ll} u^3v^3 \mid w \mid < Ax^2y^2, & u^3v^3 \mid w_u \mid < Axy^2 \\ u^3v^3 \mid w_v \mid < Ax^2y, & u^3v^3 \mid w_{uv} \mid < Axy \end{array}\right\} \text{ for } x \leqq u \leqq \pi - x,\ y \leqq v \leqq \pi - y.$$

Next, setting

$$\bar{\phi}(u, v) = \int_{kx}^{u} d\sigma \int_{ky}^{v} \phi(\sigma + x, t + y)dt,$$

and using Lemma 2, we have

$$\mid \bar{\phi}(u, v) \mid < Auv \text{ for } kx \leqq u \leqq \pi - x,\ ky \leqq v \leqq \pi - y,\ 1 \leqq k.$$

Thus

$$J_3 = \bar{\phi}(\pi - 2x, \pi - 2y)w(\pi - 2x, \pi - 2y) - \int_{kx}^{\pi-2x} \bar{\phi}(u, \pi - 2y)w_u(u, \pi - 2y)du$$

$$- \int_{ky}^{\pi-2y} \bar{\phi}(\pi - 2x, v)w_v(\pi - 2x, v)dv + \int_{kx}^{\pi-2x} du \int_{ky}^{\pi-2y} \bar{\phi}w_{uv}\, dv$$

$$= \bar{O}(x^2y^2 + y^2/k + x^2/k + 1/k^2) = \bar{o}(1).$$

This proves the lemma.

2.9. Lemma 9. *If* (L_R') *and* (L_R'') *hold, then*

$$J_4 \equiv \int_{ky}^{\pi-2y} \mid \omega(v; y) \mid dv \int_{kx}^{\pi-x} \mid \Delta_x\phi(u, v + y) \mid \frac{du}{u} = \bar{o}(1).$$

We have

$$J_4 = \bar{O}\left\{y^2\int_{ky}^{\pi-2y}\frac{dv}{v^3}\int_{kx}^{\pi-x}|\Delta_x\phi(u, v+y)|\frac{du}{u}\right\}$$

by (2.41). Thus, integrating by parts and using Lemma 6,

$$J_4 = \bar{O}\left\{y^2\beta(x, \pi; k) + y^2\int_{ky}^{\pi}\beta(x, v; k)\frac{dv}{v^4}\right\} = \bar{o}(1).$$

3.1. **Proof of Theorem I.** Because of the symmetry of the condition (L_R) with respect to the arguments of f, it is plain that Lemmas 3, 5, and 7 hold if we interchange the arguments of f in the integrals appearing there. Thus, if $0<a$, $0<b$, then

$$\int_0^{by} dv\int_{\pi-ax}^{\pi}\phi(u, v)K(u, v; x, y)du = o(1),$$

$$\int_0^{by} du\int_0^{\pi}\phi K du = o(1), \qquad \int_{\pi-ax}^{\pi} du\int_0^{\pi}\phi K\, dv = o(1).$$

Now it is readily seen by these relations and Lemmas 1, 3, 5, and 7 that

$$16S(x, y) = \left\{\int_{kx}^{\pi-2x} + 2\int_{(k+1)x}^{\pi-x} + \int_{(k+2)x}^{\pi}\right\}G(u; x, y; k)du + \bar{o}(1),$$

where

$$G = \left\{\int_{ky}^{\pi-2y} + 2\int_{(k+1)y}^{\pi-y} + \int_{(k+2)y}^{\pi}\right\}\phi K\, dv.$$

Hence, making in each of these integrals a change of variables which carries the region of integration into $(kx, ky, \pi-2x, \pi-2y)$, and collecting the terms properly, we have

$$16S = \int_{kx}^{\pi-2x}\sin\frac{\pi u}{x}du\int_{ky}^{\pi-2y}\Psi(u, v; x, y)\sin\frac{\pi v}{y}dv + \bar{o}(1),$$

where

$$\begin{aligned}\Psi = &\left\{\frac{\Delta_{x,y}\phi(u+x, v+y)}{\sin\frac{1}{2}(u+2x)\sin\frac{1}{2}(v+2y)} - \frac{\Delta_{x,y}\phi(u, v+y)}{\sin\frac{1}{2}u\sin\frac{1}{2}(v+2y)} - \frac{\Delta_{x,y}\phi(u+x, v)}{\sin\frac{1}{2}(u+2x)\sin\frac{1}{2}v}\right.\\ &\left. - \frac{\Delta_{x,y}\phi(u, v)}{\sin\frac{1}{2}u\sin\frac{1}{2}v}\right\} - \omega(v, y)\left\{\frac{\Delta_x\phi(u+x, v+y)}{\sin\frac{1}{2}(u+2x)} - \frac{\Delta_x\phi(u, v+y)}{\sin\frac{1}{2}u}\right\}\\ &- \omega(u, x)\left\{\frac{\Delta_y\phi(u+x, v+y)}{\sin\frac{1}{2}(v+2y)} - \frac{\Delta_y\phi(u+x, v)}{\sin\frac{1}{2}v}\right\}\\ &+ \omega(u; x)\omega(v; y)\phi(u+x, v+y)\\ = &\ S_1 + S_2 + S_3 + S_4, \text{ say}.\end{aligned}$$

But

$$\int_{kx}^{\pi-2x} \sin\frac{\pi u}{x} du \int_{ky}^{\pi-2y} S_1 \sin\frac{\pi v}{y} dv = \bar{O}\left\{ \int_{kx}^{\pi-2x} du \int_{ky}^{\pi-2y} | S_1 | \, dv \right\}$$
$$= \bar{O}\{\gamma(x, y; k)\} = \bar{o}(1)$$

and, by Lemma 9,

$$\int_{kx}^{\pi-2x} \sin\frac{\pi u}{x} du \int_{ky}^{\pi-2y} S_2 \sin\frac{\pi v}{y} dv$$
$$= \bar{O}\left\{ \int_{ky}^{\pi-2y} | \omega(v; y) | \, dv \int_{kx}^{\pi-2x} | \Delta_x \phi(u, v + y) | \frac{du}{u} \right\} = \bar{o}(1).$$

Similarly,

$$\int_{kx}^{\pi-2x} \sin\frac{\pi u}{x} du \int_{ky}^{\pi-2y} S_3 \sin\frac{\pi v}{y} dv = \bar{o}(1).$$

Finally,

$$\int_{kx}^{\pi-2x} \sin\frac{\pi u}{x} du \int_{ky}^{\pi-2y} S_4 \sin\frac{\pi v}{y} dv = \bar{o}(1)$$

by Lemma 8.

Thus,

$$s(x, y) = \bar{o}(1)$$

if (L_R) holds. This proves the theorem, since S is independent of k.

4.1. **Lemmas for Theorem II.** LEMMA 1. *If* (L_P') *and* (L_P'') *hold, then*

$$J_5 \equiv \int_{\pi-y}^{\pi} dv \int_{kx}^{\pi-x} \left| \Delta_x\left\{\frac{\phi}{u}\right\} \right| du = \bar{o}(1), \tag{4.11}$$

and $0<x_0<\pi$ *and* $0<k_0$ *can be found so that*

$$\eta(x, y; k) < Ay \text{ for } x \leqq x_0,\ k_0 \leqq k. \tag{4.12}$$

Further, if (L_2) *holds, then*

$$J_5(x, y; 1) = o(1). \tag{4.13}$$

The proof concerning (4.11) and (4.12) is much the same as that of Lemma 6, §2.6. Given $0<\epsilon$ we can, because of (L_P') and (L_P''), choose $1<k_0$ and x_0 so that (2.63) holds, and

$$\eta(x, y; k_0) < \epsilon y,\ \zeta(x, y; k_0) < \epsilon \text{ for } x \leqq x_0,\ y \leqq x_0. \tag{4.14}$$

Thus, if $x \leqq x_0$, $(k_0+1)z \leqq x_0 \leqq c < c+z \leqq \pi$, we have

$$\int_c^{c+z} dv \int_{k_0 z}^{\pi - z} \left| \Delta_z \left\{ \frac{\phi}{u} \right\} \right| du \leqq \pi \int_c^{c+z} \frac{dv}{v} \int_{k_0 z}^{\pi - z} \left| \Delta_z \left\{ \frac{\phi}{u} \right\} \right| du$$

$$= \pi \left\{ \int_{(k_0+1)z}^{c+z} - \int_{k_0 z}^{c} + \int_{k_0 z}^{(k_0+1)z} \right\} \frac{dv}{v} \int_{k_0 z}^{\pi - z} \left| \Delta_z \left\{ \frac{\phi}{u} \right\} \right| du$$

$$\leqq \pi \zeta(x, z; k_0) + \pi \eta \{ x, (k_0 + 1)z; k_0 \} / (k_0 z)$$

$$< \pi \epsilon + \pi \epsilon (1 + 1/k_0) < 3\pi\epsilon.$$

Choosing suitable sets of values for c and z in this inequality, and making use of (4.14) and the fact that

$$J_5(x, y; k) \leqq J_5(x, y; k_0), \quad \eta(x, y; k) \leqq \eta(x, y; k_0) \text{ for } k_0 \leqq k,$$

we deduce easily (4.11) and (4.12).

As for (4.13), we have

$$J_5(x, y; 1) = O\left[\left\{ \int_{2y}^{\pi} - \int_{y}^{\pi - y} + \int_{y}^{2y} \right\} \frac{dv}{v} \int_{x}^{\pi - x} \left| \Delta_x \left\{ \frac{\phi}{u} \right\} \right| du \right]$$

$$= O\{ \zeta_1(x, y) + \eta_1(x, 2y)/y \} = o(1)$$

by (L_2). This completes the proof.

4.2. LEMMA 2. *If* (L_P') *and* (L_P'') *hold, then* (1.61) *holds. Moreover, if* (L_2) *holds, then* (C_1^*) *holds.*

Let $\lambda(x, y; k)$ be the upper bound of

$$\xi(u, v; k)/u + J_5(x, y; k)$$

for a fixed k for $0 < u \leqq x$, $0 < v \leqq y$, and let

$$x_\mu = x\{k/(k+1)\}^\mu, \quad y_\mu = y\{k/(k+1)\}^\mu$$

for $\mu = 0, 1, \cdots$. Then, for $(k+1)x \leqq \pi$, $(k+1)y \leqq \pi$, $1 \leqq k$,

$$\phi_1^*(kx, ky) \leqq k \sum_{\mu=0}^{\infty} \left\{ y_\mu \int_0^{kx} du \int_{ky_{\mu+1}}^{ky_\mu} |\phi| \frac{dv}{v} \right\}$$

$$= k \sum_{\mu=0}^{\infty} \left[y_\mu \int_0^{kx} du \left\{ \int_{ky_{\mu+1}}^{\pi - y_{\mu+1}} - \int_{ky_\mu}^{\pi} + \int_{\pi - y_{\mu+1}}^{\pi} \right\} |\phi| \frac{dv}{v} \right]$$

$$\leqq k \sum_{\mu=0}^{\infty} \left[y_\mu \left\{ \xi(kx, y_{\mu+1}; k) + \int_0^{kx} du \int_{\pi - y}^{\pi} |\phi| \frac{dv}{v} \right\} \right]$$

$$\leqq (k+1)^3 xy\lambda(kx, y; k) + (k+1)^2 y \int_0^{kx} du \int_{\pi - y}^{\pi} |\phi| \, dv,$$

and

$$\int_0^{kx} du \int_{\pi-y}^{\pi} |\phi|\, dv \leqq k \sum_{\mu=0}^{\infty} \left\{ x_\mu \int_{\pi-y}^{\pi} dv \int_{kx_{\mu+1}}^{kx_\mu} |\phi| \frac{du}{u} \right\}$$

$$\leqq k \sum_{\mu=0}^{\infty} \left[x_\mu \left\{ J_5(x_{\mu+1}, y; k) + \int_{\pi-y}^{\pi} dv \int_{\pi-z}^{\pi} |\phi| \frac{du}{u} \right\} \right]$$

$$\leqq (k+1)^2 x\lambda(kx, y; k) + (k+1)^2 o(x).$$

Accordingly,

$$\phi_1^*(kx, ky) \leqq (k+1)^4 xy \{ 2\lambda(kx, y; k) + o(1) \}.$$

Now, if (L_P') and (L_P'') hold, then, using Lemma 1,

$$\lambda(kx, y; k) = O(1)$$

for some fixed k; while if (L_2) holds, then

$$\lambda(x, y; 1) = o(1).$$

The lemma follows.

4.3. Lemma 3. *If* (L_P') *and* (L_P'') *hold, or if* (L_R'), (L_R''), *and* (1.61) *hold, then*

$$\phi_1^*(x, y) < Axy. \tag{4.31}$$

The proof concerning (L_P') and (L_P''), as well as that concerning (L_R') and (L_R''), closely resembles the proof of (2.22). We need consider only (L_P') and (L_P''). We first observe that, by (L_P'), (L_P''), and Lemma 2, numbers $0<\epsilon$, $1<k_0$, and $0<\delta<\pi/2$ can be found such that

$$\phi_1^*(x, y) < \epsilon xy, \quad \xi(x, y; k_0) < \epsilon x, \quad \eta(x, y; k_0) < \epsilon y$$

for $x \leqq \delta$, $y \leqq \delta$. We next observe that, if $x \leqq \delta$, $(k_0+1)z \leqq \delta \leqq c < c+z \leqq \pi$, we thus have

$$\int_0^z du \int_c^{c+z} |\phi|\, dv \leqq \int_0^z du \left\{ \int_{(k_0+1)z}^{c+z} - \int_{k_0 z}^{c} + \int_{k_0 z}^{(k_0+1)z} \right\} |\phi| \frac{dv}{v}$$

$$\leqq \pi\xi(x, y; k_0) + \phi_1^*\{x, (k_0+1)z\}/(k_0 z)$$

$$< \pi\epsilon x + \epsilon x(1 + 1/k_0) < 3\pi\epsilon x.$$

Proceeding now as in the proof of (2.22), we deduce (4.31).

4.4. Lemma 4. *If* (4.31) *holds, then*

$$J_6 \equiv xy \int_{kx}^{\pi} \frac{du}{u^2} \int_{ky}^{\pi} \frac{|\phi|}{v^2}\, dv = \bar{o}(1).$$

We have, upon integrating by parts,

$$J_6 = \bar{O}\left\{xy\phi_1^*(\pi,\pi) + xy\int_{kx}^{\pi}\frac{\phi_1^*(u,\pi)}{u^3}du\right.$$

$$\left.+ xy\int_{ky}^{\pi}\frac{\phi_1^*(\pi,v)}{v^3}dv + xy\int_{kx}^{\pi}\frac{du}{u^3}\int_{ky}^{\pi}\frac{\phi_1^*}{v^3}dv\right\}$$

$$= \bar{O}(xy + y/k + x/k + 1/k^2) = \bar{o}(1).$$

4.5. Lemma 5. *If* (4.31) *holds, then* (L_R'') *is equivalent to* (L_P'').

We need consider only α and ξ. We have

$$\alpha - \xi = \bar{O}\left\{y\int_0^x du\int_{(k+1)y}^{\pi}\frac{|\phi|\,dv}{(v-y)v}\right\}$$

$$= \bar{O}\left\{y\phi_1^*(x,\pi) + y\int_{(k+1)y}^{\pi}\frac{\phi_1^*(x,v)}{(v-y)^2v}dv\right\}$$

$$= \bar{O}(xy + x/k) = \bar{o}(x);$$

and this proves the lemma.

4.6. Lemma 6. *If* (L_R') *and* (L_R'') *hold, then*

$$J_7 \equiv y\int_{ky}^{\pi}\frac{dv}{v^2}\int_{kx}^{\pi-x}|\Delta_x\phi|\frac{du}{u} = \bar{o}(1).$$

Moreover, if (L_P') *and* (L_P'') *hold, then*

$$J_8 \equiv y\int_{ky}^{\pi}\frac{dv}{v^2}\int_{kx}^{\pi-x}\left|\Delta_x\left\{\frac{\phi}{u}\right\}\right|du = \bar{o}(1).$$

We may confine our attention to the second part of this lemma. We have

$$J_8 = \bar{O}\left\{y\eta(x,\pi;k) + y\int_{ky}^{\pi}\eta(x,v;k)\frac{dv}{v^3}\right\}$$

$$= \bar{O}(y + 1/k) = \bar{o}(1)$$

by (L_P') and (L_P'') and Lemma 1.

4.7. Lemma 7. *If* (C_1^*) *holds, then* (L_R) *implies* (L_1), *and* (L_P) *implies* (L_2).

We may confine ourselves to (L_P) and (L_2). We have

$$\xi_1 = \int_0^x du\left\{\int_y^{ky} + \int_{ky}^{\pi-y}\right\}\left|\Delta_y\left\{\frac{\phi}{v}\right\}\right|dv$$

$$= \bar{O}[\phi_1^*\{x,(k+1)y\}/y + \xi(x,y;k)] = \bar{o}(x)$$

by (C_1^*) and (L_P). Thus, since ξ_1 is independent of k,

$$\xi_1 = o(x).$$

Similarly,

$$\eta_1 = o(y).$$

Accordingly, (L_2'') holds.

As for (L_2'), we have

$$\zeta_1 = \left\{ \int_x^{kx} \int_y^{\pi-y} + \int_x^{\pi-x} \int_y^{ky} - \int_x^{kx} \int_y^{ky} + \int_{kx}^{\pi-x} \int_{ky}^{\pi-y} \right\} \left| \Delta_{x,y} \left\{ \frac{\phi}{uv} \right\} \right| du\,dv$$

$$= O[\xi_1\{(k+1)x,\, y\}/x + \eta_1\{x,\, (k+1)y\}/y + \zeta] = \bar{o}(1)$$

by (C_1^*), (L_P), and (L_2''). The lemma follows.

5.1. **Proof of Theorem II.** We first note the identities

$$\frac{\Delta_{x,y}\phi}{uv} = \frac{xy}{u^2v^2}\phi + \left(1 + \frac{x}{u}\right)\frac{y}{v^2}\Delta_x\left\{\frac{\phi}{u}\right\}$$

$$+ \frac{x}{u^2}\left(1 + \frac{y}{v}\right)\Delta_y\left\{\frac{\phi}{v}\right\} + \left(1 + \frac{x}{u}\right)\left(1 + \frac{y}{v}\right)\Delta_{x,y}\left\{\frac{\phi}{uv}\right\},$$

$$\Delta_{x,y}\left\{\frac{\phi}{uv}\right\} = \frac{xy\phi}{u(u+x)v(v+y)} - \frac{y\Delta_x\phi}{(u+x)v(v+y)} - \frac{x\Delta_y\phi}{u(u+x)(v+y)}$$

$$+ \frac{\Delta_{x,y}\phi}{(u+x)(v+y)}.$$

We next note that it follows from these identities that

(5.11) $\quad \gamma \leqq J_6 + 2J_8 + 2J_9 + 4\zeta$ for $(k+1)x \leqq \pi$, $(k+1)y \leqq \pi$, $1 \leqq k$,

where

$$J_9 = x\int_{kx}^{\pi} \frac{du}{u^2} \int_{ky}^{\pi-y} \left| \Delta_y\left\{\frac{\phi}{v}\right\} \right| dv,$$

and that

(5.12) $\quad \zeta \leqq J_6 + J_7 + J_{10} + \gamma$ for $(k+1)x \leqq \pi$, $(k+1)y \leqq \pi$,

where

$$J_{10} = x\int_{kx}^{\pi} \frac{du}{u^2} \int_{ky}^{\pi-y} |\Delta_y\phi| \frac{dv}{v}.$$

Consider, then, the first part of (c). If (L_P) holds, then (4.31) holds by Lemma 3, and accordingly, by Lemma 4,

$$J_6 = \bar{o}(1).$$

Further,

$$J_8 = \bar{o}(1)$$

by Lemma 6; and plainly, by the same reasoning as in Lemma 6,

$$J_9 = \bar{o}(1).$$

Thus, by (5.11), (L_P) implies (L_R'). But, by Lemmas 3 and 5, (L_P) also implies (L_R''). The truth of the first part of (c) follows.

Now consider the second part of (c). If (L_R) and (1.61) hold, then (4.31) holds by Lemma 3, and accordingly, by Lemma 4,

$$J_6 = \bar{o}(1).$$

Further,

$$J_7 = \bar{o}(1)$$

by Lemma 6, and plainly,

$$J_{10} = \bar{o}(1).$$

Thus, by (5.12), (L_R) and (1.61) imply (L_P'). But, by Lemmas 3 and 5, (L_R) and (1.61) also imply (L_P''). The second part of (c) follows.

As for (b), the first part of (b) is trivial. The second part is proved in Lemma 7.

Turning, finally, to (a), let us first suppose that (L_1) holds. Then, plainly (L_R) holds, and thus, by (c), since (L_1) contains (C_1^*), (L_P) also holds. Accordingly, (L_2) holds by Lemma 7. Thus (L_1) implies (L_2).

Suppose, on the other hand, that (L_2) holds. Then (L_P) holds by (b), and accordingly, (L_R) holds by (c). But (L_2) also implies (C_1^*) by Lemma 2. Thus, by Lemma 7 again, (L_1) holds. This completes the proof.

6.1. **Lemmas for Theorem III.** LEMMA 1. *If (J_R) holds, then*

$$f(x, y) = O(1). \tag{6.11}$$

We choose $0 < \epsilon$ and $0 < \delta < \pi/2$ so that

$$W_1(x, y) < \epsilon,\ W_2(x, y) < \epsilon \text{ for } x \leqq 2\delta,\ y \leqq 2\delta. \tag{6.12}$$

Then we have, since $f(\delta, \delta)$ is finite,

$$\begin{aligned} |f(x, y)| &\leqq |f(x, y) - f(\delta, y)| + |f(\delta, y) - f(\delta, \delta)| + |f(\delta, \delta)| \\ &\leqq W_1(\delta, y) + W_2(\delta, \delta) + |f(\delta, \delta)| < A \end{aligned}$$

for $x \leqq \delta$, $y \leqq \delta$. This is (6.11).

6.2. LEMMA 2. *If* $0<a<b$ *and if, u being fixed, $f(u, v)$ is finite and integrable in v over $(a, b+y)$, then*

$$\int_a^b |\Delta_y f| \frac{dv}{v} \leqq \frac{y}{a} \int_a^{b+y} |d_v f(u, v)|. \tag{6.21}$$

In fact, writing

$$\psi(v) = \int_a^v |d_t f(u, t)|,$$

ψ is measurable and we have, if $\psi(b+y)$ is finite,

$$\int_a^b |\Delta_y f| \frac{dv}{v} \leqq \frac{1}{a} \int_a^b \Delta_y \psi dv \leqq \frac{1}{a} \int_b^{b+y} \psi dv \leqq \frac{y}{a} \int_a^{b+y} |d_v f|.$$

But plainly (6.21) holds if $\psi(b+y)$ is infinite; and this proves the lemma.

6.3. LEMMA 3. *If* $0<\delta<\pi$ *and if* (J_T') *holds, then*

$$J_{11} \equiv \int_\delta^{\pi-x} \frac{du}{u} \int_{ky}^{\pi-y} |\Delta_{x,y} f| \frac{dv}{v} = \bar{o}(1).$$

We have, upon applying Lemma 2,

$$J_{11} = \bar{O}\left\{ \int_0^\pi du \int_{ky}^{\pi-y} |\Delta_y f| \frac{dv}{v} \right\}$$

$$= \bar{O}\left\{ \frac{1}{k} \int_0^\pi V(u) du \right\} = \bar{o}(1).$$

6.4. LEMMA 4. *If* (J_T') *and* (J_R'') *hold, then*

$$\alpha(x, y; k) = \bar{o}(x).$$

We observe, first, that $f(+0, v)$ exists for nearly all values of v on $(0, \pi)$, for f is of bounded variation as a function of u for almost all values of v on this interval.

We observe, secondly, that $f(+0, v)$ is integrable on $(0, \pi)$. In fact, if u_0 is any number such that $0<u_0<\pi$, we have

$$|f(u, v)| \leqq |f(u_0, v)| + V(v) \text{ for } (u, v) \text{ in } Q.$$

Plainly, then, $f(+0, v)$ is the limit function of a sequence of integrable functions $\{f(u_n, v)\}$, $n=1, 2, \cdots$, satisfying a condition of the type

$$|f(u_n, v)| \leqq V_0(v) \text{ for } 0 \leqq v \leqq \pi,$$

where V_0 is integrable on $(0, \pi)$. The integrability of $f(+0, v)$ now follows from a familiar theorem of Lebesgue.*

We observe, thirdly, that

$$(6.41)\qquad \int_0^x du \int_0^\pi | f(u, v) - f(+0, v) | dv = o(x).$$

To prove this we note that

$$B(x, v) \equiv \frac{1}{x}\int_0^x | f(u, v) - f(+0, v) | du \leqq W_1(x, v).$$

Thus B tends to zero with x for nearly every v on $(0, \pi)$, and

$$B(x, v) \leqq V(v) \text{ for } 0 \leqq v \leqq \pi.$$

Accordingly, since B is integrable in v for every fixed positive value of x, (6.41) follows from the theorem of Lebesgue mentioned above.†

We observe, finally, that, upon choosing $0<\epsilon$ and $0<\delta<\pi/2$ so that (6.12) holds, we have

$$(6.42)\qquad \int_0^x du \int_{ky}^\delta | \Delta_y f | \frac{dv}{v} \leqq x\epsilon/k \text{ for } x \leqq 2\delta,\ ky \leqq \delta.$$

This results immediately upon applying Lemma 2.

The lemma now follows readily. We have

$$\alpha = \int_0^x du \int_{ky}^\delta | \Delta_y f | \frac{dv}{v} + \bar{O}\left\{ \int_0^x du \int_0^\pi | f(u, v) - f(+0, v) | dv + x\int_0^{\pi - y} | \Delta_y f(+0, v) | dv \right\} = \bar{o}(x)$$

as a consequence of (6.41), (6.42), and the well known fact that

$$\int_0^{\pi-y} | \Delta_y f(+0, v) | dv = o(1).$$

* Lebesgue, **11**, p. 375. The full theorem referred to is to the effect that, if $f_n(P)$, $n=1, 2, \cdots$, is integrable on the bounded measurable set E, if

$$|f_n(P)| < \phi(P) \text{ for } n = 1, 2, \cdots, P \text{ on } E,$$

where ϕ is integrable on E, and if

$$\lim_{n\to\infty} f_n(P)$$

exists nearly everywhere in E, then the limit function $f(P)$ of the sequence $\{f_n(P)\}$ is integrable on E, and

$$\lim_{n\to\infty} \int_E f_n(P)dP = \int_E f(P)dP.$$

† It is clear that the conclusion of this theorem remains the same if we replace the discrete variable n by a continuous one.

7.1. **Proof of Theorem III.** We first prove that (J_T) implies (L_R'). For this we choose $0<\epsilon$ and $0<\delta<\pi/2$ so that (6.12) holds; and write

$$\gamma = \left\{ \int_\delta^{\pi-x} \int_{ky}^{\delta} + \int_{kx}^{\delta} \int_\delta^{\pi-y} + \int_\delta^{\pi-x} \int_\delta^{\pi-y} + \int_{kx}^{\delta} \int_{ky}^{\delta} \right\} | \Delta_{x,y} f | \frac{du}{u} \frac{dv}{v}$$
$$= J_{11} + J_{12} + J_{13} + J_{14}, \text{ say}.$$

Now,

$$J_{11} = \bar{o}(1)$$

by (J_T') and Lemma 3, and plainly, by the same reasoning,

$$J_{12} = \bar{o}(1).$$

Further,

$$J_{13} = \bar{o}(1)$$

as is well known. It remains, then, to consider J_{14}.

We write

$$J_{14} = \int_{kx}^{t'\delta} \frac{du}{u} \int_{uy/x}^{\delta} | \Delta_{x,y} f | \frac{dv}{v} + \int_{ky}^{t''\delta} \frac{dv}{v} \int_{vx/y}^{\delta} | \Delta_{x,y} f | \frac{du}{u}$$
$$= J_{14}' + J_{14}'', \text{ say},$$

where

$$t' = \begin{cases} x/y & \text{if } x \leqq y, \\ 1 & \text{if } y < x, \end{cases} \qquad t'' = \begin{cases} 1 & \text{if } x \leqq y, \\ y/x & \text{if } y < x. \end{cases}$$

Then we have

$$J_{14}' = \bar{O}\left[\int_{kx}^{t'\delta} \frac{du}{u} \int_{uy/x}^{\delta} \{ | \Delta_y f(u, v) | + | \Delta_y f(u + x, v) | \} \frac{dv}{v} \right]$$
$$= \bar{O}\left\{ x \int_{kx}^{\infty} \frac{du}{u^2} \right\} = \bar{o}(1)$$

upon making use of (6.12) and Lemma 2. In the same way, of course, we get

$$J_{14}'' = \bar{o}(1)$$

and it follows that (J_R) implies (L_R').

The theorem is now immediate. First, since (J_T) implies (C_0) with $s=f(+0, +0)$, (J_T) implies (J_R) with $s=f(+0, +0)$. Next, because (C_1) is common to (J_R) and (L_R), it is plain, by Lemma 4 and what we have just proved, that (J_R) implies (L_R). Thus, since (J_R) implies (1.61) as a consequence of Lemma 1, it follows from Theorem II, part (c), that (J_R) implies

(L_P). Finally, making use of the fact that (J_T) implies (C_0) with $s=f(+0, +0)$ and Theorem II, parts (a) and (b), we find that (J_T) implies (L_1) with $s=f(+0, +0)$. This completes the proof.

8.1. **Lemmas for Theorem IV.** LEMMA 1. *If F satisfies (J_H), then the limits*

$$\operatorname*{limit}_{(u,v)\to(+0,+0)} F(u, v), \quad \lim_{u\to+0}\ \lim_{v\to+0} F(u, v), \quad \lim_{v\to+0}\ \lim_{u\to+0} F(u, v)$$

exist and are equal, and further, if we set

$$P + N = \int_0^u \int_0^v | d_{\sigma,t} F(\sigma, t) | + \int_0^u | d_\sigma F(\sigma, 0) | + \int_0^v | d_t F(0, t) | + | F(0, 0) |, \tag{8.11}$$

$$P - N = F \text{ for } 0 < u \leqq \pi, 0 < v \leqq \pi, \tag{8.12}$$

then both P and N satisfy the following conditions:

$$0 \leqq P < A \text{ for } 0 < u \leqq \pi, 0 < v \leqq \pi, \tag{8.13}$$

$$0 \leqq \Delta_x P \quad \text{for } 0 < u < u + x \leqq \pi, 0 < v \leqq \pi, \tag{8.14}$$

$$0 \leqq \Delta_y P \quad \text{for } 0 < u \leqq \pi, 0 < v < v + y \leqq \pi, \tag{8.15}$$

$$0 \leqq \Delta_{x,y} P \quad \text{for } 0 < u < u + x \leqq \pi, 0 < v < v + y \leqq \pi, \tag{8.16}$$

$$P \text{ is integrable in } Q. \tag{8.17}$$

The proofs of these facts, with the exception of the last, are given by Hardy.* The proof that (8.17) holds can be made to rest on a theorem of Young.† Young proves that, if the conditions (8.13) to (8.16) hold, then P is continuous at every point in the interior of Q, with the possible exception of those points found on a denumerable set of lines, each of which is parallel either to the u- or the v-axis. Thus, assuming Hardy's results, it follows that, if a is any constant, the set of points on which $P<a$ consists of an open set plus, possibly, a set of zero measure. Accordingly, P is measurable in Q; and thus, using (8.13), it follows that (8.17) holds.

8.2. LEMMA 2. *If* (8.14), (8.15), (8.16), *and* (8.17) *hold, and if*

$$0 \leqq P < Auv \text{ for } 0 < u \leqq \pi, 0 < v \leqq \pi, \tag{8.21}$$

then $P/(uv)$ satisfies (L_P') and (L_P'').

---

* Hardy, 5, pp. 57–59. Hardy defines P and N in terms of the positive and negative variation of F, but his definitions are equivalent to ours. Hardy states (8.17) without proof.

† Young, 16, p. 31.

We have

$$\zeta = \int_{kx}^{\pi-x} du \int_{ky}^{\pi-y} \left| \Delta_{x,y}\left\{\frac{P}{u^2v^2}\right\}\right| dv$$

$$= \bar{O}\left[\int_{kx}^{\pi-x} du \int_{ky}^{\pi-y} \left\{\frac{\Delta_{x,y}P}{(u+x)^2(v+y)^2} + \frac{y\Delta_x P}{(u+x)^2v^3} + \frac{x\Delta_y P}{u^3(v+y)} + \frac{xyP}{u^3v^3}\right\} dv\right]$$

$$= \bar{O}\left\{xy\int_{kx}^{\pi}\frac{du}{u^3}\int_{ky}^{\pi}\frac{P}{v^3}dv + x\int_{kx}^{\pi}\frac{du}{u^3}\int_{\pi-y}^{\pi}\frac{P}{v^2}dv + y\int_{\pi-x}^{\pi}\frac{du}{u^2}\int_{ky}^{\pi}\frac{P}{v^3}dv\right.$$

$$\left. + \int_{\pi-x}^{\pi}\frac{du}{u^2}\int_{\pi-y}^{\pi}\frac{P}{v^2}dv + \int_{kx}^{(k+1)x}\frac{du}{u^2}\int_{ky}^{(k+1)y}\frac{P}{v^2}dv\right\}$$

$$= \bar{O}\{1/k^2 + y/k + x/k + xy + 1/k^2\} = \bar{o}(1),$$

$$\xi = \bar{O}\left[\int_0^x \frac{du}{u}\int_{ky}^{\pi-y}\left\{\frac{\Delta_y P}{(v+y)^2} + \frac{yP}{v^3}\right\}dv\right]$$

$$= \bar{O}\left\{y\int_0^x\frac{du}{u}\int_{ky}^{\pi}\frac{P}{v^3}dv + \int_0^x\frac{du}{u}\int_{\pi-y}^{\pi}\frac{P}{v^2}dv\right\}$$

$$= \bar{O}(x/k + xy) = \bar{o}(x).$$

Treating η in a similar manner, we conclude the truth of the lemma.

9.1. **Proof of Theorem IV.** It is plain that (Y) implies (Y_P) and (C_1^*). As a consequence of Theorem II, then, it is enough to prove that (Y_P) implies (L_P). For this we note that, if (Y_P) holds, then the function

$$F = uv\phi$$

satisfies (J_H), for, under these circumstances, F is finitely defined everywhere in Q and

$$\int_0^{\pi}\int_0^{\pi} | d_{u,v}F | \leqq \int_0^{\pi}\int_0^{\pi} | d_{u,v}\{uvf(u,v)\} | + \pi^2 | s | < \infty,$$

$$\int_0^{\pi} | d_\sigma F(\sigma,0) | = 0, \int_0^u\int_0^v | d_t F(0,t) | = 0.$$

We note, further, that, on defining P and N as in Lemma 1, we have

$$P + N = \int_0^{\pi} | d_{u,v}(uvf) | + uv | s | < Auv \text{ for } 0 < u \leqq \pi, 0 < v \leqq \pi.$$

Applying Lemmas 1 and 2 now, we see that ϕ satisfies (L_P') and (L_P''). Since (C_1) is common to (L_P) and (Y_P), the proof is complete.

10.1. **Proof of Theorem V.** We have, if $1 < p_1$, $1 < p_2$,

$$\gamma = \bar{O}\left[\left\{\int_0^{\pi-x} du \int_0^{\pi-y} |\Delta_{x,y}f|^{p_1} dv\right\}^{1/p_1}\left\{\int_{kx}^{\infty}\frac{du}{u^{q_1}}\int_{ky}^{\infty}\frac{dv}{v^{q_1}}\right\}^{1/q_1}\right]$$

$$\left(\frac{1}{p_1}+\frac{1}{q_1}=1\right)$$

$$= \bar{O}\{x^{1/p_1}y^{1/p_1}(kx)^{1/q_1-1}(ky)^{1/q_1-1}\} = \bar{O}(k^{-2/p_1}) = \bar{o}(1),$$

$$\alpha = \bar{O}\left[\left\{\int_0^{x} du \int_0^{\pi-y} |\Delta_{y}f|^{p_2} dv\right\}^{1/p_2}\left\{\int_0^{x} du\int_{ky}^{\infty}\frac{dv}{v^{q_2}}\right\}^{1/q_2}\right]$$

$$\left(\frac{1}{p_2}+\frac{1}{q_2}=1\right)$$

$$= \bar{O}\{x^{1/p_2}y^{1/p_2}x^{1/q_2}(ky)^{1/q_2-1}\} = \bar{O}(xk^{-1/p_2}) = \bar{o}(x),$$

and

$$\gamma = \bar{O}\left\{\frac{1}{k^2xy}\int_0^{\pi-x}\int_0^{\pi-y}|\Delta_{x,y}f|\,dv\right\} = \bar{O}(1/k^2) = \bar{o}(1),$$

$$\alpha = \bar{O}\left\{\frac{1}{ky}\int_0^{x}du\int_0^{\pi-y}|\Delta_{y}f|\,dv\right\} = \bar{O}(x/k) = \bar{o}(x)$$

if $p_1=p_2=1$. Treating β in a similar manner and noting that (C_1) is common to (L_R) and (HL), we deduce the truth of the theorem.

11.1. **Proof of Theorem VI.** Since (J_H) implies (C_0) with $s=f(+0, +0)$ and f is finitely defined everywhere in Q, we have only to prove that f satisfies (1.51), (HL'), and (HL''), the last two with $p_1=p_2=p_3=1$.

Consider (1.51). We have

$$\begin{aligned}|\Delta_{x,y}(uvf)| &\leq (u+x)(v+y)|\Delta_{x,y}f| + (u+x)y|\Delta_x f| \\ &\quad + x(v+y)|\Delta_y f| + xy|f| \\ &\leq ab|\Delta_{x,y}f| + ay\left\{|\Delta_x f(u,0)| + \int_u^{u+x} d\sigma\int_0^b |d_{\sigma,t}f(\sigma,t)|\right\} \\ &\quad + xb\left\{|\Delta_y f(0,v)| + \int_0^a\int_v^{v+y}|d_{\sigma,t}f|\right\} \\ &\quad + xy \max_{0\leq\sigma\leq a,\,0\leq t\leq b} |f(\sigma,t)|\end{aligned}$$

for $0\leq u<u+x\leq a$, $0\leq v<v+y\leq b$. Thus

$$\begin{aligned}\int_0^x\int_0^y|d_{u,v}(uvf)| &\leq xy\left\{3\int_0^\pi\int_0^\pi|d_{u,v}f| + \int_0^\pi|d_u f(u,0)|\right. \\ &\quad \left. + \int_0^\pi|d_v f(0,v)| + \max_{0\leq u\leq\pi,\,0\leq v\leq\pi}|f|\right\} < Axy.\end{aligned}$$

This is (1.51).

Turning now to (HL') and (HL''), we define P and N as in Lemma 1, §8.1. Then, using the results of that lemma, we have

$$\int_0^{\pi-x} du \int_0^{\pi-y} |\Delta_{x,y}P|\, dv = O\left\{ \int_0^{\pi-x} du \int_0^{\pi-y} \Delta_{x,y}P dv \right\}$$

$$= O\left\{ \int_0^{x} du \int_0^{y} Pdv + \int_{\pi-x}^{\pi} du \int_{\pi-y}^{\pi} Pdv \right\}$$

$$= O(xy),$$

$$\int_0^{x} du \int_0^{\pi-y} |\Delta_y P|\, dv = O\left\{ \int_0^{x} du \int_{\pi-y}^{\pi} Pdv \right\} = O(xy).$$

Treating N and the other integral in (HL'') in the same manner, we infer the truth of the theorem.

12.1. **Lemmas for Theorem VII.** LEMMA 1. *If F satisfies (J_H) and is absolutely continuous* in $(a, a; \pi, \pi)$ for every $0<a<\pi$, then F coincides in the region $0<u\leqq\pi$, $0<v\leqq\pi$ with a function G which is absolutely continuous in Q.*

We first define P and N as in Lemma 1, §8.1, for $0<u\leqq\pi$, $0<v\leqq\pi$, and note that $P(u, +0)$, $P(+0, v)$, and $P(+0, +0)$ exist, the first for every u and the second for every v, on $(0, \pi)$. We complete the definition of P by setting

$$P(u, 0) = P(u, +0) \text{ for } 0 < u \leqq \pi,\ P(0, v) = P(+0, v) \text{ for } 0 < v \leqq \pi,$$
$$P(0, 0) = P(+0, +0).$$

We next note that

$$(12.11)\qquad \begin{aligned} &\int_0^{x} | d_u P(u, \pi) | = o(1), &\quad \int_0^{x}\int_0^{\pi} | a_{u,v} P | = o(1),\\ &\int_0^{y} | d_v P(\pi, v) | = o(1), &\quad \int_0^{\pi}\int_0^{y} | d_{u,v} P | = o(1). \end{aligned}$$

In fact, upon making use of (8.14) and (8.15), our definition of P on the axes, and the appropriate limiting processes, we find that

* By definition F is absolutely continuous in $(a, a; \pi, \pi)$ if (i) the functions $F(u, 0)$ and $F(0, v)$ are absolutely continuous on (a, π), and if (ii), corresponding to each $0<\epsilon$, we can so choose $0<\delta$ that, if $\{(x_i', y_i'; x_i'', y_i'')\}$, $i=1, 2, \cdots$, is any collection of rectangles contained in $(a, a; \pi, \pi)$, no two of which have a common interior point, and the total measure of which is less than δ, then
$$\textstyle\sum_{i=1}^{\infty} | f(x_i'', y_i'') - f(x_i'', y_i') - f(x_i', y_i'') + f(x_i', y_i') | < \epsilon.$$
This definition is equivalent to Carathéodory's, **2**, p. 653, but is different from Hobson's, **8**, p. 346, which requires only that (b) be satisfied.

$$0 \leqq \Delta_x P(u, \pi) \quad \text{for } 0 \leqq u < u + x \leqq \pi,$$
$$0 \leqq \Delta_{x,y} P \qquad \text{for } 0 \leqq u < u + x \leqq \pi, 0 \leqq v < v + y \leqq \pi.$$

Thus we have

$$\int_0^x \left| d_u P(u, \pi) \right| = P(x, \pi) - P(0, \pi) = o(1),$$

$$\int_0^x \int_0^\pi \left| d_{u,v} P \right| = P(x, \pi) - P(0, \pi) - P(x, 0) + P(0, 0)$$
$$= o(1) - P(x, 0) + P(0, 0) = o(1)$$

since

$$P(0, 0) = P(+0, +0) = \lim_{x \to +0} \lim_{y \to +0} P(x, y) = \lim_{x \to +0} P(x, 0).$$

This proves that the first two relations in (12.11) hold. The last two can, of course, be proved correct in a similar manner.

To complete the proof of the lemma, we now define N on the axes, as we did P, in terms of its limiting values, and we set

$$G = P - N \quad \text{for } (u, v) \text{ in } Q.$$

Then plainly N satisfies (12.11), and therefore so also does G. But

$$G = F \quad \text{for } 0 < u \leqq \pi, 0 < v \leqq \pi, \tag{12.12}$$

and therefore G is absolutely continuous in $(a, a; \pi, \pi)$ for every $0 < a < \pi$. We conclude from these two properties of G that G is absolutely continuous in Q. By (12.12), this proves the lemma.

12.2. Lemma 2. *If g is integrable over Q, then uvg satisfies (L_2).*

We have

$$\zeta_1 = \int_x^{\pi - x} du \int_y^{\pi - y} \left| \Delta_{x,y} g \right| dv = o(1),$$

$$\xi_1 = \int_0^x u\, du \int_y^{\pi - y} \left| \Delta_y g \right| dv = O\left\{ x \int_0^\pi \int_0^{\pi - y} \left| \Delta_y g \right| dv \right\} = o(x).$$

Similarly,

$$\eta_1 = o(y).$$

12.3. Lemma 3. *If $h(u)$ is integrable over $(0, \pi)$, and if $g(u, v)$ is integrable over Q, then the function $uH(u, v)$, where*

$$H(u, v) = h(u) - \int_v^\pi g(u, t)dt,$$

satisfies (L_2).

We have

$$\zeta_1 = \int_x^{\pi - x} du \int_y^{\pi - y} \left| \Delta_{x,y}\left\{\frac{H}{v}\right\}\right| dv$$

$$= O\left\{ y \int_x^{\pi-x} | \Delta_x h | du \int_y^{\pi-y} \frac{dv}{v^2} + y \int_x^{\pi-x} du \int_y^{\pi-y} \frac{dv}{v^2} \int_v^\pi | \Delta_x g(u, t) | dt \right.$$

$$\left. + \int_x^{\pi-x} du \int_y^{\pi-y} \frac{dv}{v} \int_v^{v+y} | \Delta_x g(u, t) | dt \right\}$$

$$= O\left\{ \int_0^{\pi-x} | \Delta_x h | du + \int_0^\pi du \int_0^{\pi-y} | \Delta_x g(u, t) | dt \right\} = o(1).$$

Further,

$$\xi_1 = O\left\{ y \int_0^x u\, du \int_y^{\pi-y} \frac{dv}{v^2} \int_v^\pi |g(u, t)| dt + \int_0^x u\, du \int_y^{\pi-y} \frac{dv}{v} \int_v^{v+y} | g(u, t) | dt \right\}$$

$$= O\left\{ x \int_0^x du \int_0^\pi | g(u, t) | dt \right\} = o(x).$$

Finally,

$$\eta_1 = O\left\{ \int_0^y dv \int_x^{\pi-x} | \Delta_x h | du + \int_0^y dv \int_x^{\pi-x} du \int_v^\pi | \Delta_x g(u, t) | dt \right\}$$

$$= O\left\{ y \int_0^{\pi-x} | \Delta_x h | du + y \int_x^{\pi-x} du \int_0^\pi | \Delta_x g(u, t) | dt \right\} = o(y).$$

This completes the proof.

13.1. **Proof of Theorem VII.** By our hypothesis, F can be so defined on the axes as to satisfy (J_H). Moreover, F obviously is absolutely continuous in $(a, a; \pi, \pi)$ for every $0 < a < \pi$. Thus, by Lemma 1, F coincides in the region $0 < u \leqq \pi$, $0 < v \leqq \pi$ with a function G absolutely continuous in Q.

Now, since G is absolutely continuous in Q, there exist functions $g(u, v)$, $h(u)$, $l(v)$, the first integrable over Q and the last two over $(0, \pi)$, such that, for (u, v) in Q, G is given by*

---

* See Hobson, 8, pp. 592, 615, or Carathéodory, 2, p. 654.

$$G(u, v) = \int_u^\pi d\sigma \int_v^\pi g(\sigma, t)dt - \int_u^\pi h(\sigma)d\sigma - \int_v^\pi l(t)dt + G(\pi, \pi)$$
$$= g_1 - h_1 - l_1 + G(\pi, \pi), \text{ say.}$$

We proceed to express f in terms of g, h, and l.

We have, for $0<u<u+x\leqq\pi$, $0<v<v+y\leqq\pi$,

$$\Delta_{x,y}\left\{ \int_0^u d\sigma \int_0^v f dt\right\} = \Delta_{x,y}(uvG)$$
$$= (u + x)(v + y)\Delta_{x,y}g_1 + (u + x)y\Delta_x G + x(v + y)\Delta_y G + xyF.$$

Hence, upon dividing each member of this equation by xy, setting $y=x$, and letting x tend to zero, we have*

$$f(u, v) = uvg + uH + vL + F, \tag{13.11}$$

where

$$H(u, v) = h(u) - \int_v^\pi g(u, t)dt,$$

$$L(u, v) = l(v) - \int_u^\pi g(\sigma, v)d\sigma,$$

for almost all (u, v) in Q.

The theorem now follows. Since g and h are integrable, uvg and uH satisfy (L_2) by Lemmas 2 and 3. Similarly, since l is integrable, vL satisfies (L_2). Applying Theorem II, then, we see that (L_1) holds with f replaced by $uvg+uH+vL$ and s, by zero. On the other hand, since F satisfies (J_H), it satisfies (Y) with $s=F(+0, +0)$ by Theorem VI. Hence, by Theorem IV, F satisfies (L_1) with $s=F(+0, +0)$. Combining these results with (13.11), we reach the desired conclusion.

References

1. Adams, C. R., to be published.

2. Carathéodory, C., *Vorlesungen über Reelle Funktionen*, Leipzig and Berlin, 1927.

3. Merriman, G. M., *On certain theorems regarding summable series and their application to double and triple Fourier series*, American Journal of Mathematics, vol. 47 (1925), pp. 125–139.

4. Gergen, J. J., *Convergence and summability criteria for Fourier series*, Quarterly Journal of Mathematics, Oxford Series, vol. 1 (1930), pp. 252–275.

5. Hardy, G. H., *On double Fourier series*, Quarterly Journal of Pure and Applied Mathematics, vol. 37 (1905), pp. 53–79.

6. ——— *On certain convergence criteria for Fourier series*, Messenger of Mathematics, vol. 49 (1919–20), pp. 149–155.

* The proof concerning the double differences can be found in Hobson, **8**, p. 614, or in Carathéodory, **2**, p. 496; that concerning the simple differences, in Hobson, **8**, pp. 588, 611, or in Carathéodory, **2**, p. 658.

7. Hardy, G. H. and Littlewood, J. E., *Solution of the Cesàro summability problem for power series and Fourier series*, Mathematische Zeitschrift, vol. 19 (1923), pp. 67–96.

8. Hobson, E. W., *Theory of Functions of a Real Variable*, Cambridge, vol. 1, 1927.

9. ——— *Theory of Functions of a Real Variable*, Cambridge, vol. 2, 1926.

10. Küstermann, W. W., *Über Fouriersche Doppelreihen*, Inaugural Dissertation, München, 1913.

11. Lebesgue, H., *Sur l'intégration des fonctions discontinues*, Annales Scientifiques de l'École Normale Supérieure, (3), vol. 27 (1910), pp. 361–450.

12. Pringsheim, A., *Elementare Theorie der unendlichen Doppelreihen*, Sitzungsberichte der Mathematisch-Physikalischen Classe der Akademie der Wissenschaften zu München, vol. 27 (1897), pp. 101–153.

13. Tonelli, L., *Serie Trigonometriche*, Bologna, 1928.

14. ——— *Sulla Convergenza delle Serie Doppie di Fourier*, Annali di Matematica, (4), vol. 4 (1927), pp. 29–72.

15. Young, W. H., *Multiple Fourier series*, Proceedings of the London Mathematical Society, vol. 11 (1913), pp. 133–184.

16. ——— *On multiple integrals*, Proceedings of the Royal Society of London, vol. 93 (1916–17), pp. 28–42.

HARVARD UNIVERSITY,
CAMBRIDGE MASS.

ON THE NUMERATORS OF THE CONVERGENTS OF THE STIELTJES CONTINUED FRACTIONS*

BY

JACOB SHERMAN

Introduction. The object of this paper is the study of the numerators of the infinite continued fractions introduced by Stieltjes [1, 2].† They are of two types:

(a) the "associated" continued fraction:

$$K(z) \equiv \frac{\lambda_1 \mid}{\mid z - c_1} - \frac{\lambda_2 \mid}{\mid z - c_2} - \frac{\lambda_3 \mid}{\mid z - c_3} - \cdots, \tag{1}$$

the nth convergent of which will be denoted by $\Omega_n(z)/\phi_n(z) \equiv K_n(z)$ $(n=0, 1, 2, 3, \cdots)$;

(b) the "corresponding" continued fraction:

$$W(z) \equiv \frac{b_1|}{|z} - \frac{b_2|}{|1} - \frac{b_3|}{|z} - \frac{b_4|}{|1} - \cdots, \tag{2}$$

the nth convergent of which will be denoted by $U_n(z)/V_n(z) \equiv W_n(z)$ $(n=0, 1, 2, 3, \cdots)$. In (a) λ_i, c_i are real constants with $\lambda_i > 0$ for $i=1, 2, 3, \cdots$; $\Omega_n(z)$, $\phi_n(z)$ are polynomials of degree $n-1$ and n respectively.‡ In (b) b_i are real constants, $b_1 > 0$, $b_{2i+1}b_{2i} > 0$ for $i=1, 2, \cdots$. $U_{2n+\epsilon}(z)$, $V_{2n+\epsilon}(z)$ are polynomials of degree $(n+\epsilon-1)$ and $(n+\epsilon)$ respectively, where $\epsilon = 0, 1$. For our study we shall need certain results from the theory of continued fractions [2] and of the so-called moments problem [3].

(i) *The convergence of the associated and corresponding continued fractions.* Here of fundamental importance is Grommer's Selection Theorem [2]:

From every sequence of convergents of a continued fraction of type (1) *or* (2) *there may be selected a sub-sequence, which for all non-real z converges to a Stieltjes integral of the form* $\int_{-\infty}^{\infty}(1/(z-u))d\psi(u)$.

Here $\psi(z)$ (and hereafter $\psi_1(z)$, $\psi_2(z)$, $\cdots$) denotes generally a bounded monotonic non-decreasing function with infinitely many points of increase in $(-\infty, \infty)$, such that all integrals $\int_{-\infty}^{\infty} z^n d\psi(z) \equiv \alpha_n$ ("moments") exist (for

* Presented to the Society, March 26, 1932; received by the editors April 28, 1932.

The author wishes to express his gratitude to Professor J. Shohat for many valuable suggestions concerning the subject of this paper.

† The numbers in brackets refer to the list of literature at the end of the paper.

‡ We note that $\Omega_n(z) \equiv \lambda_1 z^{n-1} + \cdots$ does not depend on λ_2 and $\phi_n(z) \equiv z^n + \cdots$ does not depend on λ_1. This will be useful in our later discussion.

$n=0, 1, 2, 3, \cdots$), with $\alpha_0>0$. We may without loss of generality assume $\psi(-\infty)=0$. The continued fraction (1) or (2) is said to be "*associated*" with or "*corresponding*" to, respectively, the integral $\int_{-\infty}^{\infty}(1/(z-u))d\psi(u)$ or its formal development

$$P\left(\frac{1}{z}\right) = \sum_{n=0}^{\infty} \frac{\alpha_n}{z^{n+1}}.$$

In symbols:

$$F(z) \equiv \int_{-\infty}^{\infty} \frac{d\psi(u)}{z-u} = P\left(\frac{1}{z}\right) \equiv \sum_{n=0}^{\infty} \frac{\alpha_n}{z^{n+1}} \sim \frac{\lambda_1 \mid}{\mid z - c_1} - \frac{\lambda_2 \mid}{\mid z - c_2} - \cdots \tag{3}$$
$$\left(\int_{-\infty}^{\infty} z^n d\psi(z) = \alpha_n\right).$$

The association (3) means formally [3]

$$\int_{-\infty}^{\infty} \frac{d\psi(u)}{z-u} - K_n(z) = \frac{\alpha'}{z^{2n+1}} + \frac{\alpha''}{z^{2n+2}} + \cdots; \tag{4}$$

similarly the correspondence means formally

$$\int_{-\infty}^{\infty} \frac{d\psi(u)}{z-u} - W_n(z) = \frac{\beta'}{z^{n+1}} + \frac{\beta''}{z^{n+2}} + \cdots.$$

It is known that a continued fraction of form (1) may be obtained from one of form (2) by "contraction," and then

$$A_{2n}(z) \equiv \Omega_n(z), \quad B_{2n}(z) \equiv \phi_n(z) \quad (n = 0, 1, 2, 3, \cdots).$$

The "association" (4) shows [2] that the $\{\phi_n(z)\}$* constitute an orthogonal set of polynomials with regard to the distribution $d\psi(x)$, i.e.

$$\int_{-\infty}^{\infty} \phi_m(x)\phi_n(x)d\psi(x) = \begin{cases} 0, & m \neq n, \\ > 0, & m = n \end{cases} \quad (m, n = 0, 1, 2, 3, \cdots), \tag{5}$$

where $(-\infty, \infty)$ may be "reducible" (see page 70) to a sub-interval (a, b). We might have $\psi(x)=\int_{-\infty}^{x}p(x)dx$ ($p(x)\geqq 0$ in $(-\infty, \infty)$); then in all our formulas $d\psi(x)$ is to be replaced by $p(x)dx$. In particular (5) becomes

$$\int_{-\infty}^{\infty} \phi_m(x)\phi_n(x)p(x)dx = 0, \; m \neq n \quad (m, n = 0, 1, 2, 3, \cdots).$$

Here we say that $\{\phi_n(x)\} \equiv \{\phi_n(x; a, b; p)\}$ forms an orthogonal set corresponding to the *characteristic function* $p(x)$ ($p(x)\geqq 0$) in $(-\infty, \infty)$). In some

* Throughout this paper z represents the complex variable $x+iy$.

instances we might have $\psi(x)=\psi(a)$ for $x<a$ and $\psi(x)=\psi(b)$ for $x>b$ ($p(x)\equiv 0$ outside (a, b)) in which case $\int_{-\infty}^{\infty}(1/(z-u))d\psi(u)$ reduces to $\int_a^b(1/(z-u))d\psi(u)$ and (5) becomes

$$\int_a^b \phi_m(x)\phi_n(x)d\psi(x) = 0, \quad m \neq n \quad (m, n = 0, 1, 2, 3, \cdots).$$

The function $F(z)\equiv\int_{-\infty}^{\infty}(1/(z-u))d\psi(u)$ in (3) is known [2] to be regular and analytic in any finite closed region of the complex z-plane which does not contain any portion of the real axis. Such a region we call, for brevity, an "Ω-region" [3]. In such a region $1/F(z)$ is also analytic. In fact it is readily proved that $F(z)$ has no zeros in an Ω-region. We now introduce, following Hamburger [3], the continued fraction

$$K_n(z, t) \equiv \frac{\lambda_1}{|z - c_1} - \frac{\lambda_2}{|z - c_2} - \cdots - \frac{\lambda_{n-1}}{|z - c_{n-1}} - \frac{\lambda_n}{|z - c_n + t}, \tag{6}$$

λ_i, c_i as in (1), t real, arbitrary; $K_n(z, \infty)\equiv K_{n-1}(z)$. Evidently $K_n(z, 0) \equiv K_n(z)$.

We take, from Hamburger [3], the following

DEFINITION. *The continued fraction $K(z)$ converges "completely" to the function $F(z)$ of the type* (3) *if, for arbitrarily small $\epsilon>0$ and for every Ω-region, $|K_n(z, t)-F(z)|<\epsilon$ (z in Ω, $n\geqq N$) for every real arbitrary t, including $t=\infty$, where N depends on ϵ and Ω only.*

(ii) *The $[\alpha_n]_0^\infty$-moments problem.* By this is meant the determination of $\psi(z)$ of the above nature, given the real set $[\alpha_n]$ ($n=0, 1, 2, \cdots$) of its moments. If the set determines $\psi(z)$ uniquely (disregarding additive constants) the moments problem is said to be "determined"; otherwise it is said to be "indetermined." In case of an "association," as given in (4), we say that $\psi(z)$ is the solution of the moments problem related to $P(1/z)$ or $K(z)$, or, see (5), to the orthogonal set $\{\phi_n(x)\}$. Hamburger [3] has shown that the complete convergence of the associated continued fraction $K(z)$ to the integral $\int_{-\infty}^{\infty}(1/(z-u))d\psi(u)$ is a necessary and sufficient condition for the moments problem related to $K(z)$ to be determined over the interval $(-\infty, \infty)$.

Our purpose is the study of the polynomials $\{U_{2n}(x)\}$, $\{U_{2n+1}(x)\}$, $\{V_{2n+1}(x)\}$ and their relations to the orthogonal polynomials $\{\phi_n(x)\}$. Some of the results concerning $\{U_{2n}(x)\}$ have been presented by J. Shohat and J. Sherman in [4].

1. The orthogonality properties of the numerators $\Omega_n(x)$. The study of the $\Omega_n(x)$ is based upon the continued fraction

$$K'(z) \equiv \frac{\lambda_2 \,|}{|z - c_2} - \frac{\lambda_3 \,|}{|z - c_3} - \cdots \tag{7}$$

(λ_i, c_i the same as in (1)), whose successive convergents will be denoted by $P_n(z)/Q_n(z) \equiv K_n'(z)$ $(n=0, 1, 2, 3, \cdots)$.

LEMMA 1. $\Omega_{n+1}(z) = \lambda_1 Q_n(z)$ $(n \geqq 0)$.

We have for $K(z)$, $K'(z)$ respectively

$$\Omega_{n+1}(z) = (z - c_{n+1})\Omega_n(z) - \lambda_{n+1}\Omega_{n-1}(z) \qquad (n = 1, 2, 3, \cdots) \tag{8}$$
$$(\Omega_0(z) \equiv 0\,,\ \Omega_1(z) = \lambda_1,\ \Omega_2(z) = \lambda_1(z - c_2));$$

$$Q_n(z) = (z - c_{n+1})Q_{n-1}(z) - \lambda_{n+1}Q_{n-2}(z) \qquad (n = 2, 3, 4, \cdots) \tag{9}$$
$$(Q_0(z) \equiv 1,\ Q_1(z) = z - c_2).$$

The truth of the lemma is seen by comparing (8) and (9). In fact, they represent the same difference equation with the initial conditions $Q_0(z)$, $Q_1(z)$ and $\Omega_1(z)$, $\Omega_2(z)$, respectively, differing only by a constant factor (λ_1). Our lemma leads to the following important conclusion:

$K'(z)$ is an infinite continued fraction, the denominators of whose convergents are, to within a constant factor independent of n, identical with the numerators of the convergents of $K(z)$.

LEMMA 2. *$K'(z)$ is associated with one and only one "positive definite" power series.**

In symbols

$$K'(z) \sim P_1(1/z) \equiv \sum_{n=0}^{\infty} \frac{\beta_n}{z^{n+1}}\,.$$

This follows directly from the results due to Hamburger [3], since, in $K'(z)$, all $\lambda_i > 0$ $(i=1, 2, 3, \cdots)$.

LEMMA 3. *If $K(z)$ converges in a certain Ω-region, then $K'(z)$ converges in that same region to a function which is regular and analytic in that region.*

---

* The series

$$P(1/z) \equiv \sum_{n=0}^{\infty} \frac{\alpha_n}{z^{n+1}}$$

is said to be "positive definite" if all the determinants

$$\Delta_m \equiv [\alpha_{i,j}]_{i,j=0}^{m-1} > 0 \text{ for } m=1, 2, \cdots,$$

where $\alpha_{i,j} = \alpha_{i+j}$. We set also $\Delta_0 \equiv 1$.

It is known [2] that, if $K(z)$ is convergent, $K'(z)$ must either converge or diverge to infinity. The latter is impossible, for, by Grommer's Theorem, we may select a sub-sequence $\{K_{n_i}'(z)\}$ $(i=1, 2, 3, \cdots)$ of the convergents of $K'(z)$ which in the Ω-region converges to a Stieltjes integral of the form (3):

$$F_1(z) \equiv \int_{-\infty}^{\infty} \frac{d\psi_1(u)}{z-u},$$

and $F_1(z)$ is regular and analytic in Ω. Hence $K'(z)$ itself converges in that Ω-region to $F_1(z)$.

COROLLARY. *If $F(z)$ and $F_1(z)$, both analytic and non-vanishing in Ω, are the limit functions of $K(z)$ and $K'(z)$ respectively, then*

$$F(z) = \frac{\lambda_1}{z - c_1 - F_1(z)} \qquad (z \text{ in } \Omega). \tag{10}$$

LEMMA 4. *If $K(z)$ converges completely to*

$$F(z) \equiv \int_{-\infty}^{\infty} \frac{d\psi(u)}{z-u},$$

then $K'(z)$ converges completely to a certain

$$F_1(z) \equiv \int_{-\infty}^{\infty} \frac{d\psi_1(u)}{z-u}.$$

By the definition of $K_n(z, t)$ (see (6)) we write

$$\begin{gathered} K_n(z,t) = \frac{\lambda_1}{z - c_1 - K_{n-1}'(z,t)} \\ \left(K_{n-1}'(z,t) \equiv \frac{\lambda_2 \mid}{\mid z - c_2} - \frac{\lambda_3 \mid}{\mid z - c_3} - \cdots - \frac{\lambda_{n-1} \mid}{\mid z - c_{n-1}} - \frac{\lambda_n \mid}{\mid z - c_n + t}\right). \end{gathered} \tag{11}$$

Hence $K_{n-1}'(z, t)$ plays the same rôle for $K'(z)$ as $K_{n-1}(z, t)$ does for $K(z)$. Using the definition of complete convergence,

$$| K_n(z,t) - F(z) | < \epsilon \qquad (n \geqq N(\epsilon, \Omega), z \text{ in } \Omega), \tag{12}$$

we are lead through (10) and (11), and with the same $z, n, \epsilon, t, \Omega$ as in (12), to

$$| K_{n-1}'(z,t) - F_1(z) | \leqq \frac{\epsilon}{\lambda_1} \cdot \left| \frac{\lambda_1}{K_n(z,t)} \right| \cdot \left| \frac{\lambda_1}{F(z)} \right|.$$

Furthermore, we know that

(a) $$\lambda_1 > 0; \ |F(z)| \geqq h > 0 \qquad (z \text{ in } \Omega),$$

(b) $$|K_n(z, t)| \geqq \frac{h}{2} > 0 \qquad (z \text{ in } \Omega, n \geqq N) \ (\text{by } (12)).$$

Hence

$$\frac{1}{\lambda_1}\left|\frac{\lambda_1}{K_n(z, t)}\right| \cdot \left|\frac{\lambda_1}{F(z)}\right|$$

is bounded above for all real t (including $t = \infty$) and for all z in Ω, and thus our lemma is proved.

Combining our lemmas, we are in a position to prove

THEOREM I. *The numerators* $\{\Omega_n(x)\}$ *of the continued fraction* $K(x)$ *or, which is the same, the numerators of the even convergents of the continued fraction* $W(x)$, *constitute a set of polynomials orthogonal with regard to a function* $\psi_1(x)$ *of the same type as* $\psi(x)$, *i.e.*

(13) $$\int_{-\infty}^{\infty} \Omega_m(x)\Omega_n(x) d\psi_1(x) = 0, \ m \neq n \qquad (m, n = 1, 2, 3, \cdots).$$

THEOREM II. $\psi_1(z)$, *as a solution of the moments problem associated with* $K'(z)$, *is uniquely determined if* $\psi(z)$ *is uniquely determined as a solution of the moments problem associated with* $K(z)$.

The proof of Theorem I follows from the fact that $K'(z)$, being of the same type as $K(z)$, is, therefore, associated, in the sense of (4), with at least one Stieltjes integral of type (3): $\int_{-\infty}^{\infty}(1/(z-u))d\psi_1(u)$, which leads to orthogonality relations (13) similar to (5), and here also $(-\infty, \infty)$ may reduce to a certain sub-interval.

Theorem II follows from Lemma 4 combined with Hamburger's necessary and sufficient condition for the determined character of the moments problem as given above (see page 66).

We can also state

THEOREM III. *Let* (λ, L) *and* (λ', L') *denote the "true" intervals of orthogonality for* $\{\phi_n(x)\}$ *and* $\{\Omega_n(x)\}$ *respectively.*

(i) $(\lambda', L') \subset (\lambda, L)$.

(ii) *If* (λ, L) *is finite, then, in general,* $\lambda' = \lambda$ *and* $L' = L$. *Any sub-interval of* (λ, L) *which does not contain an interval of constancy of* $\psi(x)$ *possesses the same property with regard to* $\psi_1(x)$.

(iii) *If* (λ, L) *is infinite, so is* (λ', L'); *more precisely, if* $(\lambda, L) \equiv (\lambda, \infty)$, $(-\infty, L)$ $(\lambda, L$ *finite) then, respectively,* $L' = \infty$ *or* $\lambda' = -\infty$. *If* $(\lambda, L) \equiv (-\infty, \infty)$, *then* $(\lambda', L') \equiv (-\infty, \infty)$ [4].

By a "non-reducible" or "true" interval of orthogonality we mean [4] the interval determined by the limits of the least and greatest roots of $\phi_n(x)$ as n increases indefinitely. Stieltjes [1] proved, for the interval $(0, \infty)$, to which we can always reduce the intervals (λ, ∞), $(-\infty, L)$ (λ or L finite), that, if the largest root of $\phi_n(x)$ approaches infinity with n, then an infinite sequence of such roots approach infinity. Since we know that the zeros of $\phi_n(x)$ separate those of $\Omega_n(x)$, the theorem is established for the simply infinite interval. We need thus consider only the case of $(\lambda, L) \equiv (-\infty, \infty)$. For this purpose we make use of the following results of Hamburger [3]: $\psi_n(x)$ is defined as a weight function of order n relative to the positive definite series

$$P\left(\frac{1}{x}\right) \equiv \sum_{n=0}^{\infty} \frac{\alpha_n}{x^{n+1}} \sim \int_{-\infty}^{\infty} \frac{d\psi(y)}{x - y}$$

if

(a) $\psi_n(x)$ is a step function with exactly n points of increase, $x_{n,i}$. Let the saltus at such a point equal $M_{n,i}(i = 1, 2, \cdots, n)$. Thus

$$\begin{aligned} &\psi_n(x) = 0\,(-\infty \leqq x < x_{n,1}), && \psi_n(x) = M_{n,1}\,(x_{n,1} \leqq x < x_{n,2}), \\ &\psi_n(x) = M_{n,2}\,(x_{n,2} \leqq x < x_{n,3}), && \psi_n(x) = \alpha_0\,(x_{n,n} \leqq x \leqq +\infty). \end{aligned} \tag{14}$$

(b) At least the first $(2n-1)$ moments of the weight function $\psi_n(x)$ are identical with those of $\psi(x)$:

$$\int_{-\infty}^{\infty} x^{\nu} d\psi_n(x) = \alpha_\nu \quad (\nu = 0, 1, 2, \cdots, 2n - 1).$$

From the theory of continued fractions [1] we know that the zeros of $\phi_{n+1}(x)$ separate those of $\phi_n(x)$, i.e., if the zeros of $\phi_n(x)$, in order of magnitude, are denoted by $x_{n,1}, x_{n,2}, \cdots, x_{n,n}$ we have

$$x_{n+1,1} < x_{n,1} < x_{n+1,2} < x_{n,2} < \cdots < x_{n,n-1} < x_{n+1,n} < x_{n,n} < x_{n+1,n+1}.$$

This shows that the largest zero and the next largest zero of $\phi_n(x)$ increase with n. Since we are dealing with a true interval of orthogonality, we have, by hypothesis,

$$x_{n,n} \to +\infty,\ x_{n,1} \to -\infty \qquad (n \to \infty). \tag{A}$$

Let us assume

$$x_{n,n-1} \to L', \qquad x_{n,2} \to \qquad (n \to \infty;\ \lambda', \text{ or } L', \text{ or both, finite}). \tag{B}$$

We then prove that assumption (B) contradicts the hypothesis (A).

Consider the relations [2]

$$\frac{\Omega_n(x)}{\phi_n(x)} = \sum_{i=1}^{n} \frac{M_{n,i}}{x - x_{n,i}}$$

$$\left(\sum_{i=1}^{n} M_{n,i}x_{n,i}^{\nu} = \alpha_\nu;\ \nu = 0, 1, 2, \cdots, 2n - 1;\ M_{n,i} > 0\right). \tag{15}$$

In particular

$$\sum_{i=1}^{n} M_{n,i}x_{n,i}^{2} = \alpha_2 = \int_{-\infty}^{\infty} x^2 d\psi(x) > 0.$$

It follows that

$$M_{n,n}x_{n,n}^{2} < \alpha_2,\ M_{n,n} \to 0\ \ (n \to \infty)\ \ (\text{by (A)}).$$

In the relations which define $\psi_n(x)$, let us take for $M_{n,i}$ and $x_{n,i}$ the same quantities as in (15). Then, as was shown by Hamburger [3], (b) is satisfied. Furthermore, there is a sub-sequence $\{\psi_{n_\nu}(x)\}$ $(\nu = 1, 2, \cdots)$ which, for $\nu \to \infty$, approaches $\psi(x)$ as a limit at all its points of continuity. For brevity we shall write $\psi_\nu(x)$ in place of $\psi_{n_\nu}(x)$. Then

$$\psi_\nu(x) = \sum_{i=1}^{\nu} M_{\nu,i} = \alpha_0 \qquad (x_{\nu,\nu} \leqq x \leqq +\infty).$$

Take $x = L''$, any point of continuity of $\psi(x)$, such that $x_{\nu,\nu-1} < L' < L'' < x_{\nu,\nu}$. Then

$$\psi_\nu(L'') = \sum_{i=1}^{\nu'} M_{\nu,i} = \alpha_0 - M_{\nu,\nu}$$

and

$$\lim_{\nu=\infty} \psi_\nu(L'') = \psi(L'') = \alpha_0 - \lim_{\nu=\infty} M_{\nu,\nu} = \alpha_0.$$

Therefore, since $\psi(\infty) = \alpha_0$, $\psi(x)$ is constant in the interval (L', ∞) and all integrals involving $d\psi(x)$ would have L' for the upper limit of integration. Hence, $x_{\nu,\nu} < L'$, for all ν, in contradiction to (A).*

In similar manner, using $M_{\nu,1}$ instead of $M_{\nu,\nu}$, we can show that the assumption $\lim_{n=\infty} x_{n,2} = \lambda'$, λ' finite, involves a contradiction ($\psi(x) = 0$ in $(-\infty, \lambda')$). This proves our theorem for, if $\lim x_{n,n-1} = \lim x_{n,n} = \infty$ and $\lim x_{n,2} = \lim x_{n,1} = -\infty$ $(n \to \infty)$, then, due to separation, the greatest and the least zeros of $\Omega_n(x)$ approach $+\infty$ and $-\infty$, respectively. Henceforth we shall, in general, denote by (a, b) the true interval of orthogonality, finite or infinite, for the set $\{\phi_n(x)\}$.

* It is known that, if

$$\int_a^b \phi_m(x)\phi_n(x)d\psi(x) = 0,\ m \neq n \qquad (m, n \neq 0, 1, 2, \cdots)$$

then all zeros of $\phi_n(x)$ $(n = 1, 2, \cdots)$ are real, distinct and between a and b.

2. We now turn to the corresponding continued fraction $W(z)$ as given in (2) and its convergents $U_n(z)/V_n(z)$. Hereafter we assume, with Stieltjes [1], that $b_i>0$ $(i=1, 2, \cdots)$. Then the corresponding Stieltjes integral is of the form

$$\int_0^\infty \frac{d\psi(u)}{z-u} = \sum_{i=0}^\infty \frac{\alpha_i}{z^{i+1}} \sim W(z) \qquad \left(\alpha_i = \int_0^\infty x^i d\psi(x)\right) \tag{16}$$

($(0, \infty)$ perhaps being reducible to (a, b) with $a>0$ [1, 2]). Since we get the continued fraction (1) by contraction from (2), $U_{2n}(z) \equiv \Omega_n(z)$, $V_{2n}(z) \equiv \phi_n(z)$ $(n=0, 1, 2, \cdots)$, we shall restrict ourselves to the odd convergents $U_{2n-1}(z)/V_{2n-1}(z)$ $(n \geqq 1)$. From the difference relations

$$U_{2n-1}(z) = zU_{2n-2}(z) - b_{2n-1}U_{2n-3}(z),$$
$$U_{2n-2}(z) = U_{2n-3}(z) - b_{2n-2}U_{2n-4}(z) \qquad (n \geqq 2,\ U_0 \equiv 0),$$

with similar expressions for $V_{2n-1}(z)$, $V_{2n-2}(z)$, we derive easily

$$U_{2n-1}(z) = [z - (b_{2n-1} + b_{2n-2})]U_{2n-3}(z) - b_{2n-2}b_{2n-3}U_{2n-5}(z) \tag{17}$$
$$[n > 2,\ U_1(z) = b_1;\ U_3(z) = b_1(z - b_3)],$$

$$V_{2n-1}(z) = [z - (b_{2n-1} + b_{2n-2})]V_{2n-3}(z) - b_{2n-2}b_{2n-3}V_{2n-5}(z) \tag{18}$$
$$[n > 2;\ V_1(z) = z;\ V_3(z) = z(z - b_2 - b_3)].$$

Since $U_{2n-1}(z)$ and $V_{2n-1}(z)$ are polynomials of degree $(n-1)$ and n, respectively, with $V_{2n-1}(0)=0$, let

$$U_{2n-1}(z) = Q_{n-1}(z), \quad V_{2n-1}(z) = zS_{n-1}(z) \qquad (n \geqq 1) \tag{19}$$

($Q_{n-1}(z)$, $S_{n-1}(z)$ are polynomials of degree $n-1$). The difference equations (17) and (18) then lead to the following continued fractions of type (1):

$$K''(z) \equiv \frac{b_2\ |}{|z - b_3} - \frac{b_3b_4\ |}{|z - b_4 - b_5} - \frac{b_5b_6\ |}{|z - b_6 - b_7} - \cdots, \tag{20}$$

with convergents $P_n(z)/Q_n(z)$,

$$K'''(z) \equiv \frac{b_1b_2\ |}{|z - b_2 - b_3} - \frac{b_3b_4\ |}{|z - b_4 - b_5} - \frac{b_5b_6\ |}{|z - b_6 - b_7} - \cdots, \tag{21}$$

with convergents $R_n(z)/T_n(z)$.

Consider first $K'''(z)$. Compare $S_n(z)$ as computed from (18) and (19) with $T_n(z)$ as computed from (21) (making $S_0(z)=T_0(z)=1$). We see that

$$S_n(z) = T_n(z) \qquad (n = 1, 2, \cdots). \tag{22}$$

Moreover, all the $b_i (i=1, 2, \cdots)$ being positive, $K'''(z)$ may be formally associated with a positive definite series, $P_1(1/z)\sum_{i=0}^{\infty}\beta_i/z^{i+1}$ (in the sense of (3)) the coefficients, β_i, of which may be computed through known formulas [2], and also with a Stieltjes integral $\int_0^\infty (1/(z-u))d\psi_2(u)$ of type (16), so that

$$\int_0^\infty \frac{d\psi_2(u)}{z-u} \equiv \sum_{i=0}^{\infty} \frac{\beta_i}{z^{i+1}} \sim K'''(z) \qquad \left(\beta_i = \int_0^\infty x^i d\psi_2(x)\right).$$

Hence, by the fundamental property of association (see (4), (5)),

$$\int_0^\infty T_m(x)T_n(x)d\psi_2(x) = 0 \qquad (m \neq n;\, m, n = 0, 1, 2, \cdots).$$

Let

$$P\left(\frac{1}{z}\right) = \sum_{i=0}^{\infty} \frac{\alpha_i}{z^{i+1}} \tag{23}$$

be the series associated with $K(z)$, as given in (1), which, as was stated, is obtained by contraction from $W(z)$ in (2). Hence, since [5]

$$\lambda_n = b_{2n-2}b_{2n-1},\ c_n = b_{2n+1} + b_{2n}\ (n \geqq 2); \qquad \lambda_1 = b_1,\ c_1 = b_2, \tag{24}$$

$K(z)$ can be re-written as follows:

$$K(z) \equiv \frac{b_1 \mid}{\mid z - b_2} - \frac{b_2 b_3 \mid}{\mid z - b_3 - b_4} - \frac{b_4 b_5 \mid}{\mid z - b_5 - b_6} - \cdots .$$

We have then for the α_i [2]

$$\alpha_0 = \lambda_1 = b_1,\ \alpha_1 = \lambda_1 c_1 = b_1 b_2,\ \alpha_2 = \lambda_1(\lambda_2 + c_1^2) = b_1 b_2(b_2 + b_3), \cdots .$$

In a similar manner we can express the β_i related to the continued fraction $K'''(z)$. A simple comparison of the values of α_i and β_i thus obtained gives the relation

$$\beta_i = \alpha_{i+1} \qquad (i = 0, 1, 2, \cdots). \tag{25}$$

But, according to Stieltjes,

$$K(z) = \lim_{n=\infty} \frac{\Omega_n(z)}{\phi_n(z)} = \lim_{n=\infty} \frac{A_{2n}(z)}{B_{2n}(z)} = \int_0^\infty \frac{d\psi(u)}{z-u}, \quad z \text{ not in } (0, \infty),$$

$$\alpha_i = \int_0^\infty x^i d\psi(x) \qquad (i = 0, 1, 2, \cdots),$$

and (25) shows that a solution of the $[\beta_i]_0^\infty$-moments problem is $\psi_2(x)$, determined by the relation

$$d\psi_2(x) = x\,d\psi(x).$$

Combining (19) and (22) we get the following result:*

$$\int_0^\infty \phi_m(x)\phi_n(x)d\psi(x) = 0 \text{ implies } \int_0^\infty T_m(x)T_n(x)x d\psi(x) = 0$$
$$(m \neq n;\ m, n = 0, 1, 2, \cdots).$$

We now turn to $K''(z)$, as given in (20). The very form of (20) shows, by reasoning similar to that employed above, that

$$\int_0^\infty Q_m(x)Q_n(x)d\psi_3(x) = 0,\ m \neq n \qquad (m, n = 0, 1, 2, \cdots). \tag{26}$$

Comparison of $K'(x)$ and $K''(x)$ (see (7), (20), (24) combined with what we know about the distribution of the roots of $\Omega_n(x)$) shows that

$$\int_0^\infty \Omega_m(x)\Omega_n(x)d\psi_1(x) = 0,\ m \neq n \qquad (m, n = 0, 1, 2, \cdots). \tag{26a}$$

Furthermore, if we compare the coefficients of the power series of type (23), associated with $K'(z)$ and $K''(z)$ respectively, we find again, as above, that we may take in (26) and (26a)

$$x d\psi_3(x) = d\psi_1(x),$$

i.e. the relations (26) may be rewritten as

$$\int_0^\infty Q_m(x)Q_n(x)\frac{1}{x}d\psi_1(x) = 0, \quad m \neq n \quad (m, n = 0, 1, 2, \cdots).$$

Let, further,

$$K^{\text{iv}}(z) \equiv \frac{b_3b_4 \mid}{\mid z - b_4 - b_5} - \frac{b_5b_6 \mid}{\mid z - b_6 - b_7} - \cdots.$$

Then, following the method used in Lemma 4, it can be shown that if $K^{\text{iv}}(z)$ is completely convergent, so is $K''(z)$, and that the complete convergence of $K(z)$ implies that of $K'''(z)$, which in turn implies the complete convergence of $K^{\text{iv}}(z)$. Hence, the complete convergence of $K(z)$ implies that of $K''(z)$. The discussion of the intervals of orthogonality is similar to that of the previous case. We may, then, combine our results into the general

THEOREM IV. *A Stieltjes continued fraction of type* (2)

$$W(z) = \frac{b_1 \mid}{\mid z} - \frac{b_2 \mid}{\mid 1} - \frac{b_3 \mid}{\mid z} - \frac{b_4 \mid}{\mid 1} - \cdots \qquad (b_i > 0;\ i = 1, 2, 3, \cdots)$$

---

* This result was first obtained, in an entirely different manner, by J. Shohat [6].

with the convergents $U_i(z)/V_i(z)$ $(i=0, 1, 2, \cdots)$ *gives rise to four sets of orthogonal polynomials of degree* $0, 1, 2, \cdots$:

$$\text{(a)} \qquad \int_0^\infty V_{2m}(x)V_{2n}(x)d\psi(x) = 0,$$

$$\text{(b)} \qquad \int_0^\infty U_{2m}(x)U_{2n}(x)d\psi_1(x) = 0,$$

$$\text{(c)} \qquad \int_0^\infty S_m(x)S_n(x)x\,d\psi(x) = 0 \qquad \left(S_n(x) \equiv \frac{V_{2n+1}(x)}{x},\ n \geqq 1;\ S_0(x) = 1\right),$$

$$\text{(d)} \qquad \int_0^\infty Q_m(x)Q_n(x)\frac{1}{x}d\psi_1(x) = 0 \qquad (Q_n(x) \equiv U_{2n+1}(x))$$

$$(m \neq n;\ m, n = 0, 1, 2, 3, \cdots).$$

If the moments problem related to the set $\{\phi_n(x)\} \equiv \{V_{2n}(x)\}$ *is determined then the moments problems related to the other three sets are also determined.*

We notice that the main feature of our proof consisted in constructing continued fractions of the type $K(z)$ for which each of the sets of polynomials under discussion are the denominators of the successive convergents.

3. **Differential equation for $\Omega_n(x)$.** If $d\psi(x)=p(x)dx$, and $p(x)$ is of the form

$$(27) \qquad p(x) = e^{Q(x)}\prod_{i=1}^{s}(x-a_i)^{A_i} \qquad (A_i > -1;\ a_i = \text{const.};\ Q(x) \text{ a polynomial})$$

then it has been shown by J. Shohat [7] that

$$(28) \qquad F'(x) = T(x)F(x) + R(x) \qquad \left(F(x) = \int_a^b \frac{p(y)}{x-y}dy\right)$$

$(T(x)=p'(x)/p(x)$, $R(x)$ a rational function), and that the corresponding orthogonal polynomials satisfy a homogeneous linear differential equation of second order

$$(29) \qquad A_n(x)y_n''(x) + B_n(x)y_n'(x) + C_n(x)y_n(x) = 0$$

$$(A_n(x),\ B_n(x),\ C_n(x) \text{ polynomials}).$$

In particular, for the classical orthogonal polynomials, of Jacobi, Laguerre and Hermite, $p(x)$ is of type (27) and (29) assumes a very simple form. In fact $A_n(x)$, $B_n(x)$ are polynomials independent of n, of degree $\leqq 2, 1$, respectively, while C_n is a constant depending on n only. They are as follows:

(30)

	$p(x)$	interval of orthogonality	$A_n(x) \equiv A(x)$	$B_n(x) \equiv B(x)$	$C_n(x) \equiv C_n$
Jacobi (J)	$(x-a)^{\alpha-1}(b-x)^{\beta-1}$ $(\alpha, \beta > 0)$	(a, b) finite	$(x-a)(b-x)$	$\alpha b - \beta a - (\alpha+\beta)x$	$n(\alpha+\beta+n-1)$
Laguerre (L)	$e^{-x}x^{\alpha-1}$ $(\alpha > 0)$	$(0, \infty)$	x	$\alpha - x$	n
Hermite (H)	e^{-x^2}	$(-\infty, \infty)$	1	$-2x$	$2n$

Furthermore,

$$p(x) = (1/A(x)) \exp \int \frac{B(x)}{A(x)} dx. \tag{31}$$

In order to derive a differential equation for $\Omega_n(x)$ we need one more relation from the theory of continued fractions [2]:

$$\phi_n(x)F(x) = \Omega_n(x) + R_n(x) \qquad \left(R_n(x) = \frac{\alpha'}{x^{n+1}} + \frac{\alpha''}{x^{n+2}} + \cdots, \alpha' \neq 0\right). \tag{32}$$

Moreover, [7], for $p(x)$ of the type (27), $R_n(x)/p(x)$ is a second solution of the differential equation (29). From (32) we find

$$R_n(x)/p(x) = \phi_n(x)F(x)/p(x) - \Omega_n(x)/p(x).$$

Substituting this expression into (29), we find the desired equation after a somewhat lengthy but simple computation:

$$\begin{aligned} &A_n(x)\Omega_n''(x) + [B_n(x) - 2A_n(x)p'(x)/p(x)]\Omega_n'(x) + [C_n(x) - B_n(x)p'(x)/p(x) \\ &\quad - A_n(x)p''(x)/p(x) + 2A_n(x)p'(x)^2/p(x)^2]\Omega_n(x) \\ &= [A_n(x)R'(x) - (A_n(x)p'(x)/p(x))R(x) + B_n(x)R(x)]\phi_n(x) \\ &\quad + 2A_n(x)R(x)\phi_n'(x). \end{aligned} \tag{33}$$

(33) can be very greatly simplified for the classical polynomials (30). We show that here $R(x)$ has a very simple expression:

$$R(x) = (A''(x) - B'(x) + C_1)/A(x) \equiv k_1/A(x) \qquad \text{(see (36) below)}. \tag{34}$$

Using the expression for $p(x)$ from (31), (33) becomes

$$\begin{aligned} &A(x)\Omega_n''(x) + [2A'(x) - B(x)]\Omega_n'(x) + [A''(x) - B'(x) + C_n]\Omega_n(x) \\ &\qquad = [A(x)R(x)]'\phi_n(x) + 2A(x)R(x)\phi_n'(x) \qquad (n = 1, 2, \cdots). \end{aligned} \tag{35}$$

We now let $n = 1$, and observe that $\Omega_n(x)$ is a polynomial of degree $n-1$, and that $A(x)$ and $B(x)$ do not depend on n. Making the leading coefficient of $\Omega_n(x)$ equal 1 we obtain

$$(A''(x) - B'(x) + C_1) = [A(x)R(x)]'(x - c_1) + 2A(x)R(x).$$

Let (see (30))

$$A''(x) - B'(x) + C_n = k_n \text{ (constant independent of } x). \tag{36}$$

We have

$$[A(x)R(x)]'(x - c_1) + 2A(x)R(x) = k_1,$$

and this we treat as a differential equation in $A(x)R(x)$, which gives

$$A(x)R(x) = \frac{D}{(x - c_1)^2} + k_1/2 \quad (D = \text{const.}). \tag{37}$$

It is now easy to show that $D=0$. In fact, substituting (37) into (35), we get, for $n=2$, on the right side

$$\begin{aligned} &- 2D/(x - c_1)^3\phi_2(x) + 2[D/(x - c_1)^2 + k_1/2]\phi_2'(x) \\ &= k_1\phi_2'(x) - [2D/(x - c_1)^3][\phi_2(x) - (x - c_1)\phi_2'(x)], \end{aligned}$$

and this can reduce to a polynomial (as in the left side of (35)), if and only if $D=0$.

We thus proved (34), which, substituted into (35), gives the following differential equation for $\Omega_n(x)$ in the classical cases:

$$\begin{aligned} A(x)\Omega_n''(x) + [2A'(x) - B(x)]\Omega_n'(x) + k_n\Omega_n(x) = k_1\phi_n'(x) \\ (k_n = A''(x) - B'(x) + C_n;\ n = 1, 2, \cdots). \end{aligned} \tag{38}$$

The coefficients of the differential equation for the numerators are thus seen to be the coefficients of the adjoint of the differential equation (29) for the denominators.

We shall write equation (35) explicitly (see 30):

(39) $\mathrm{J}: (x - a)(b - x)\Omega_n''(x) + [2(a + b) - \alpha b - \beta a + (\alpha + \beta - 4)x]\Omega_n'(x)$
$+ [(\alpha + \beta)(n + 1) + n(n - 1) - 2]\Omega_n(x) = 2(\alpha + \beta - 1)\phi_n'(x)[8].$

(40) $\mathrm{L}: x\Omega_n''(x) - (\alpha - 2 - x)\Omega_n'(x) + (n + 1)\Omega_n(x) = 2\phi_n'(x).$

(41) $\mathrm{H}: \Omega_n''(x) + 2x\Omega_n'(x) + 2(n + 1)\Omega_n(x) = 4\phi_n'(x).$

Making use of the simple expression of $R(x)$ as given in (34), we can find another expression for $F(x)$ from (28):

$$F(x) = p(x)\left[\frac{k_1}{2}\int^x \frac{dx}{A(x)p(x)} + C\right] \quad (C = \text{const.}). \tag{42}$$

Remarks. (i) If we differentiate (38) twice and, where $y_n \equiv \phi_n(x)$, (29) once, eliminate $\phi_n{}^i(x)$ $(i=0, 1, 2, 3)$, we obtain for $\Omega_n(x)$ a linear homogenous equation of the 4th order.

(ii) The general solution of (38) is

$$\Omega_n(x) + D_1 p(x)\phi_n(x) + D_2 R_n(x) \qquad (D_{1,2} = \text{const.})$$

for [7] $p(x)\phi_n(x)$ and $R_n(x)$ are solutions of

$$A(x)y_n''(x) + [2A'(x) - B(x)]y_n'(x) + [A''(x) - B'(x) + C_n]y_n(x) = 0.$$

4. **The differential equations for $\{\Omega_n(x)\}$ of Laguerre and Hermite cases as limiting cases of that for the $\{\Omega_n(x)\}$ of the Jacobi case.** Following a method indicated by P. Appell and J. Kampé de Fériet [9], we write the differential equations for the $\{\Omega_n(x)\}$ for the Jacobi case in the interval (0, 1) and the characteristic function $x^{\alpha-1}(1-x)^{\beta-1}$ or $\Omega_n[x;\, 0,\, 1;\, x^{\alpha-1}(1-x)^{\beta-1}]$:

$$\begin{aligned} x(1-x)\Omega_n''(x) &+ [2-\alpha+(\alpha+\beta-4)x]\Omega_n'(x) \\ &+ [(\alpha+\beta)(n+1)+n(n-1)-2]\Omega_n(x) \\ &= 2(\alpha+\beta-1)n\phi_{n-1}[x;\,0,\,1;\,x^{\alpha}(1-x)^{\beta}]; \end{aligned} \tag{43}$$

since in the classical cases generally $\phi_n'[x;a,b;p] = n\phi_{n-1}'[x;a,b;Ap]$ where $A \equiv A(x)$ is the coefficient of $\phi_n''(x)$ in (29). In (43) let $\beta-1=s$, $x=x_1/s$, divide by $s^{1/2}\,s^{(1-\alpha/2)}\,(\alpha+s-3)$ and then let $s\to\infty$. We thus obtain

$$x_1\Omega_n''(x_1) + (2-\alpha-x_1)\Omega_n'(x_1) + (n+1)\Omega_n(x_1) = 2\phi_n'(x_1)$$

which is the differential equation (40) for the $\Omega_n(x_1)$ in the Laguerre case.

Again, write the differential equation for $\Omega_n[x;\,-1,1;\,(1+x)^{\alpha-1}(1-x)^{\beta-1}]$

$$\begin{aligned} (1-x^2)\Omega_n''(x) &+ [-\alpha+\beta+(\alpha+\beta-4)x]\Omega_n'(x) \\ &+ [(\alpha+\beta)(n+1)+n(n-1)-2]\Omega_n(x) = 2(\alpha+\beta-1)\phi_n'(x); \end{aligned}$$

let $\alpha-1=\beta-1=s$, $x=x_1 s^{-1/2}$, divide through by $s^{1/4}(s-1)$, and let $s\to\infty$:

$$\Omega_n''(x_1) + 2x_1\Omega_n'(x_1) + 2(n+1)\Omega_n(x_1) = 4\phi_n'(x_1),$$

which is the differential equation (41) for the $\Omega_n(x_1)$ in the Hermite case.

Hence, the numerators as well as the denominators of the associated continued fractions of Laguerre and Hermite cases may be considered as limiting cases of those of the Jacobi case.

5. **Relations between $\{\phi_n(x)\}$ and $\{\Omega_n(x)\}$.** (i) It is evident, from what has been said before, that we may write the associated continued fraction $K'(x)$, having $\{\Omega_n(x)\}$ as denominators of its convergents, directly from the given $K(x)$. Also, given the corresponding continued fraction $W(x)$, we may, from (20) and (21), write the continued fractions of type $K(x)$ where $U_{2n}(x)$, $U_{2n+1}(x)$, $(1/x)\,V_{2n+1}(x)$ will play the same rôle as $V_{2n}(x)$. Let

$$\phi_n(x) \equiv x^n - S_n x^{n-1} + \cdots;$$
$$\Omega_n(x) \equiv x^{n-1} - \sigma_{n-1}x^{n-2} + \cdots \qquad (n = 1, 2, \cdots).$$

We know [5] that $S_n = \sum_{i=1}^{n} c_i$ and since $K'(x)$ is obtained by raising the indices of the λ_i's and c_i's by 1, it follows directly that

$$\sigma_n = S_{n+1} + c_1 \qquad (n = 1, 2, \cdots). \tag{44}$$

Let a_n, ρ_n $(n=0, 1, \cdots)$ represent the "normalizing factors" for the sets $\{\phi_n(x)\}$, $\{\Omega_n(x)\}$ respectively, i.e.

$$\int_{-\infty}^{\infty} \varphi_m(x)\varphi_n(x)d\psi(x) = \int_{-\infty}^{\infty} \omega_m(x)\omega_n(x)d\psi_1(x) = \delta_{m,n} \quad (m, n = 0, 1, 2, \cdots),$$

$$\varphi_n(x) \equiv a_n\phi_n(x), \quad \omega_n(x) \equiv \rho_n\Omega_{n+1}(x).$$

We know [5] that

$$a_n^2 = (\lambda_1\lambda_2 \cdots \lambda_{n+1})^{-1} \qquad (n = 0, 1, 2, \cdots).$$

It follows then, in the same manner as for (44), that

$$\rho_n = \lambda_1^{1/2} a_{n+1} \qquad (n = 0, 1, \cdots)$$

if we choose the first partial numerator in $K'(x)$ equal to λ_2, which can be done without any loss of generality.

(ii) *Relations between the moments of the distributions $d\psi(x)$ and $d\psi_1(x)$.* Introducing proper constant factors, we can choose $\psi(x)$, $\psi_1(x)$ in (3), (13), so that $\alpha_0 = \int_{-\infty}^{\infty} d\psi(x) = 1$; $\beta_0 = \int_{-\infty}^{\infty} d\psi_1(x) = \lambda_2$. The "associations"

$$F(x) \equiv P(1/x) = \sum_{i=0}^{\infty} \frac{\alpha_i}{x^{i+1}} \sim K(x) \qquad \left(\alpha_i = \int_{-\infty}^{\infty} x^i d\psi(x),\ \alpha_0 = 1\right),$$

$$F_1(x) \equiv P_1(1/x) = \sum_{i=0}^{\infty} \frac{\beta_i}{x^{i+1}} \sim K'(x) \qquad \left(\beta_i = \int_{-\infty}^{\infty} x^i d\psi_1(x),\ \beta_0 = \lambda_2\right)$$

lead to the formal relation $F(x) = (x - c_1 - F_1(x)^{-1}$ from which we have the following relations between the α_i and β_i:

$$-\beta_i = \begin{vmatrix} 0 & 0 & & \cdots 1 & \alpha_1 \\ 0 & \cdot & & \cdots \alpha_1 & \alpha_2 \\ \cdot & \cdot & \cdot & \cdots & \alpha_3 \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ 1 & \alpha_1 & \alpha_2 & \cdots \alpha_i & \alpha_{i+1} \\ \alpha_1 & \alpha_2 & & \cdots \alpha_{i+1} & \alpha_{i+2} \end{vmatrix} \qquad (i \geqq 1).$$

(iii) *The symmetric case.* Let the $\{\phi_n(x)\}$ be a set of "symmetric" orthogonal polynomials, i.e.

$$S_n = c_n = 0, \quad \phi_n(x) \equiv (-1)^n \phi_n(-x) \qquad (n = 1, 2, \cdots).$$

Then from the similarity of the difference equations satisfied by $\{\phi_n(x)\}$ and $\{\Omega_n(x)\}$ (see (8), (9)), we conclude that the $\{\Omega_n(x)\}$ also form a symmetric set. However, as may be surmised from the fact that the set $\{\phi_n(x)\}$ involves one more essential constant (c_1) than does the set $\{\Omega_n(x)\}$, the symmetry of the set $\{\phi_n(x)\}$ is not necessary for that of the set $\{\Omega_n(x)\}$. In fact, if we take the Jacobi case in the interval $(-1, 1)$, then, from the general formula [10]

$$c_n = (\alpha - \beta)(\alpha + \beta - 2)/[(\alpha + \beta + 2n - 2)(\alpha + \beta + 2n - 4)]$$
$$(n > 1;\ c_1 = (\alpha - \beta)/(\alpha + \beta)),$$

we see that, if $\alpha+\beta=2$, $\alpha\neq\beta$, then $c_n=0$ $(n=2, 3, \cdots)$, but $c_1\neq 0$. Thus, in this case, the set $\{\Omega_n(x)\}$ is symmetric while the set $\{\phi_n(x)\}$ is not.

6. **Some particular classical cases.** In this section we shall study some particular classical cases in which we obtain an explicit expression for $p_1(x)$, the characteristic function for the set $\{\Omega_n(x)\}$, also some simple relations between $\{\phi_n(x)\}$ and $\{\Omega_n(x)\}$.

(i) *The sets* $\{\phi_n(x)\}$ *and* $\{\Omega_n(x)\}$ *are identical (disregarding constant factors), i.e.*

$$\phi_n(x) \equiv \Omega_{n+1}(x) \qquad (n = 0, 1, 2, \cdots). \tag{45}$$

For the sake of simplicity we consider the case of symmetric $\{\phi_n(x)\}$. Our hypothesis leads to

$$F(x) = D_1/(x - D_2F(x)) \quad (D_{1,2} = \text{const.}).$$

Solving this equation for $F(x)$ and choosing a proper sign for the radical and proper value for the constants $D_{1,2}$,* we obtain

$$F(x) = \pi/(x + (x^2 - 1)^{1/2}) = \int_{-1}^{1} \frac{(1 - y^2)^{1/2}}{x - y} dy \ [10].$$

* The formula

$$F(x) = \int_{-\infty}^{\infty} \frac{p(y)dy}{x - y}$$

shows that for $|x|\to\infty$, $F(x)\to 0$. On the other hand, in the continued fraction

$$\frac{D_1|}{|x} - \frac{D_2|}{|x} - \cdots,$$

the $\phi_n(x)$ do not depend on D_1, the $\Omega_n(x)$ do not depend on D_2 and contain D_1 as a factor. Hence having D_2 we can choose D_1 so as to have $D_1D_2=1$.

Hence

$$(a, b) = (-1, 1); \quad \psi(x) = \int_{-1}^{x} (1 - x^2)^{1/2} dx, \quad p(x) = (1 - x^2)^{1/2},$$

and the above is the only case in which, under the given conditions, (45) is satisfied. The explicit expression for $\Omega_n(x)$ is $(\sin n\theta)/\sin\theta$ $(x = \cos\theta)$.†

(ii) $\phi_n'(x) = \Omega_n(x)$ $(n = 1, 2, \cdots)$ *(within a constant factor)*. The differential equation (38) for the $\Omega_n(x)$ where we substitute now $\phi_n'(x)/n = \Omega_n(x)$ gives

$$A(x)\Omega_n''(x) + (2A'(x) - B(x))\Omega_n'(x) + (C_n - C_1)\Omega_n(x) = 0. \tag{46}$$

Applying (31) to (46), we obtain

$$p_1(x) = 1/p(x).$$

But we know that the characteristic function for $\phi_n'(x)$ is $A(x)p(x)$.

Hence,

$$p_1(x) = 1/p(x) = A(x)p(x); \quad p(x) = A(x)^{-1/2}. \tag{47}$$

In view of (30), the only case when (47) holds is the Jacobi case for $\alpha = \beta = 1/2$ where the $\phi_n(x)$ are the so-called "trigonometric polynomials" $\cos n \operatorname{arc} \cos x$ and

$$\Omega_n(x) = \frac{\sin n\theta}{\sin\theta} \qquad (x = \cos\theta).$$

The characteristic [10] function for the $\{\phi_n(x)\}$ is, in the interval $(-1, 1)$, $(1-x^2)^{-1/2}$ and the characteristic function for the $\{\Omega_n(x)\}$ is, by (47), $(1-x^2)^{1/2}$.

It is interesting to note that the $\Omega_n(x)$ so obtained are the same as in (i), for $p_1(x)$ obtained from (47) is identical with that found in (i).

(iii) $\Omega_n(x)$ *of Jacobi case, in the interval* $(-1, 1)$ *with* $\alpha + \beta = 1$, *i.e. with* $p(x) = (1+x)^{\alpha-1}(1-x)^{-\alpha}$ $(1 > \alpha > 0)$. The differential equation (39) for $\Omega_n(x)$ of the Jacobi case in the interval $(-1, 1)$, with arbitrary $\alpha, \beta > 0$, is

$$\begin{aligned}(1 - x^2)\Omega_n''(x) &+ [-\alpha + \beta + (\alpha + \beta - 4)x]\Omega_n'(x) \\ &+ [(\alpha + \beta)(n + 1) + n(n - 1) - 2]\Omega_n(x) = 2(\alpha + \beta - 1)\phi_n'(x).\end{aligned} \tag{48}$$

† Here, disregarding constant factors,

$$\frac{\Omega_n(\cos\theta)}{\phi_n(\cos\theta)} = \frac{\sin n\theta}{\sin(n+1)\theta} = \frac{(x+(x^2-1)^{1/2})^n - (x-(x^2-1)^{1/2})^n}{(x+(x^2-1)^{1/2})^{n+1} - (x-(x^2-1)^{1/2})^{n+1}}.$$

If we take $|x| > 1$ and let $n \to \infty$ we get in the limit $(x+(x^2-1)^{1/2})^{-1} = F(x)/\pi$ in accordance with Markoff's theorem.

The case $\alpha+\beta=1$ deserves special attention, for then (48) becomes a homogenous differential equation of the same type as that for $\phi_n(x)$:

$$(1-x^2)\Omega_n''(x)+[1-2\alpha-3x]\Omega_n'(x)+(n+1)(n-1)\Omega_n(x)=0 \qquad (n=1,2,\cdots). \tag{49}$$

Comparing (49) with (30) (where n, α, β are replaced by $n-1$, $2-\alpha$, $\alpha+1$ respectively) we conclude that the $\Omega_n(x)$ are identical with the $\phi_n(x)$ corresponding to $(1+x)^{1-\alpha}(1-x)^{\alpha}=1/p(x)$. For the interval (0, 1), making use of Theorem IV, we thus obtain

THEOREM V. *If in the continued fraction* $W(x)$, *given by* (2), *with the convergents* $U_i(x)/V_i(x)$ $(i=0, 1, \cdots)$, *the set* $\{V_{2n}(x)\}$ *is orthogonal in* (0, 1) *with the characteristic function* $x^{\alpha-1}(1-x)^{-\alpha}$ $(1>\alpha>0)$, *then the other three sets*, $\{(1/x)\ V_{2n+1}(x)\}$, $\{U_{2n}(x)\}$, $\{U_{2n+1}(x)\}$, *are orthogonal in* (0, 1) *with the characteristic functions* $x^{\alpha}(1-x)^{-\alpha}$, $x^{1-\alpha}(1-x)^{\alpha}$, $x^{-\alpha}(1-x)^{\alpha}$, *respectively.*

The trigonometric polynomials considered above in (ii) are evidently a special case $\alpha=\beta=1/2$.

(iv) Using the difference equations ((17), (18)) and the fact that $K'''(x)$ has, for denominators of its convergents, polynomials orthogonal with respect to the distribution $xd\psi(x)$ $(b_1=1)$, we obtain

$$V_{2n}(xd\psi(x))+U_{2n}(xd\psi(x))=U_{2n+1}(d\psi(x)) \qquad (b_1=1;\ n=1,2,\cdots). \tag{50}$$

In fact, $T_i(x)+R_i(x)=Q_i(x)$ $(i=0, 1)$, and $T_n(x)$ and $R_n(x)$ satisfy the same difference equation, except for initial conditions, as $Q(x)$; then

$$R_n(x)+T_n(x)=Q_n(x) \qquad (n=1,2,\cdots).$$

7. **Expansion of $\{\Omega_n(x)\}$ in terms of $\{\phi_n(x)\}$.** Our starting point is the integral representation of $\Omega_n(x)$ [2],

$$\Omega_n(x)=\int_a^b \frac{\phi_n(x)-\phi_n(y)}{x-y}\,d\psi(y).$$

We find easily, writing

$$\frac{\phi_n(x)-\phi_n(y)}{x-y}=L_{n,n-1}(y)\phi_{n-1}(x)+L_{n,n-2}(y)\phi_{n-2}(x)+\cdots,$$

$$\begin{aligned} L_{n,n-1}(y)&=1,\\ L_{n,n-2}(y)&=y-c_n,\\ L_{n,n-3}(y)&=(y-c_{n-1})L_{n,n-2}(y)-\lambda_nL_{n,n-1}(y),\\ L_{n,n-4}(y)&=(y-c_{n-2})L_{n,n-3}(y)-\lambda_{n-1}L_{n,n-2}(y),\\ &\cdots\cdots\cdots\cdots\cdots\cdots \end{aligned} \tag{51}$$

We may thus exhibit $L_{n,n-i-1}(y)$ as the denominators of the ith convergent of the continued fraction

$$\frac{1\ |}{|y-c_n} - \frac{\lambda_n\ |}{|y-c_{n-1}} - \frac{\lambda_{n-1}\ |}{|y-c_{n-2}} - \cdots - \frac{\lambda_3\ |}{|y-c_2}, \tag{51a}$$

whose relation to $K(y)$ is obvious. These denominators may also be expressed as determinants [2]. We have

$$L_{n,0} = \begin{vmatrix} y-c_n & \lambda_n & 0 & \cdot & \cdots & \cdot \\ 1 & y-c_{n-1} & \lambda_{n-1} & 0 & \cdots & \cdot \\ 0 & 1 & \cdot & \cdot & \cdots & \cdot \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & 0 & 1 & y-c_3 & \lambda_3 \\ \cdots & \cdots & \cdot & 0 & 1 & y-c_2 \end{vmatrix} \qquad (n = 2, 3, \cdots),$$

and $L_{n,n-i}(y)$ is formed from $L_{n,0}(y)$ by dropping the $n-i$ last rows and columns.

Our expansion is then

$$\Omega_n(x) = \sum_{i=1}^{n} A_{n,n-i}\phi_{n-i}(x),$$

$$A_{n,n-i} = \int_a^b \Omega_n(x)\phi_{n-i}(x)d\psi(x) = \int_a^b L_{n,n-i}(x)d\psi(x).$$

Another method for obtaining the expansion under consideration is furnished by the differential equation (38). This method, first applied by Christoffel [8] to the Legendre case, may be epitomized as follows. Substitute in (38)

$$\phi_n'(x) = \sum_{i=1}^{n} g_{n,n-i}\phi_{n-i}(x), \tag{52}$$

$$\Omega_n(x) = \sum_{i=1}^{n} A_{n,n-i}\phi_{n-i}(x) \qquad (g_{n,n-i},\ A_{n,n-i} = \text{const.}), \tag{53}$$

and equate the coefficients of $\phi_{n-i}(x)$ (for an expansion of a polynomial in terms of $\phi_i(x)$ is necessarily unique). This leads to linear relations involving the $g_{n,n-i}$ and the $A_{n,n-i}$. These become especially simple in the cases which follow.

(i) *Hermite case*: $p(x)=e^{-x^2}$. This is the simplest case, since $\phi_n'(x)\equiv n\phi_{n-1}(x)$, i.e. $g_{n,n-1}=n$, $g_{n,n-2}=g_{n,n-3}=\cdots=g_{n,0}=0$. Carrying out the indicated computation, we obtain

$$\Omega_n(x) = \sum_{i=0} A_{n,n-(2m+1)}\phi_{n-(2m+1)}(x),$$

(54) $$A_{n,n-(2m+1)} = [(-1)^m/2^m](n-m-1)(n-m-2) \cdot \cdots (n-2m) \qquad \left(m = 1, 2, \cdots, \left[\frac{n-1}{2}\right]\right).$$

(ii) *Laguerre case*: $p(x) = e^{-x}$. If we write out the explicit expression of Laguerre polynomials

$$\phi_n(x) = x^n - n^2x^{n-1} + \frac{n^2(n-2)^2}{2!}x^{n-2} + \cdots \quad (n = 0, 1, \cdots),$$

we readily obtain the coefficients in the expansion (52) by direct computation:

$$g_{n,n-i} = (-1)^{i+1}n!/(n-i)! \qquad (n = 1, 2, \cdots; i = 1, 2, \cdots, n).$$

Using this result we proceed as in the previous case:

$$\Omega_n(x) = \sum_{i=1}^{n} A_{n,n-i}\phi_{n-i}(x);$$

(55) $$\begin{aligned} A_{n,n-1} &= 1, \\ A_{n,n-2} &= -2(n-1), \\ A_{n,n-3} &= \frac{2(n-2)}{2n-2}\left[\frac{n!}{(n-2)!} - (n-2)A_{n,n-2}\right], \\ A_{n,n-4} &= -\frac{2(n-3)}{2n-3}\left[\frac{n!}{(n-3)!} + (n-3)A_{n,n-3}\right], \\ &\cdots\cdots\cdots\cdots\cdots\cdots \end{aligned}$$

8. **Relation between $\psi(x)$ and $\psi_1(x)$.** We assume that $K(z)$, hence $K'(z)$, are related to determined moments problems whose solutions are $\psi(x)$ and $\psi_1(x)$ respectively. In order to establish a relation between these two functions we make use of the following formula (Perron [2], p. 372):

$$\frac{\psi(x-0)+\psi(x+0)}{2} - \frac{\psi(x_0-0)+\psi(x_0+0)}{2} = \lim_{y=+0} R\left\{\frac{1}{\pi i}\int_{x_0+iy}^{x+iy} F(z)dz\right\}$$
$$(z = x+iy,\ a < x_0,\ x < b),$$

where the path of integration is a straight line parallel to the x-axis. If we apply the same formula to $\psi_1(x)$ and express the result in terms of $\psi(x)$, using (10), we obtain, since

$$\lim_{y=+0} R\left\{\frac{1}{\pi i}\int_{x_0+iy}^{x+iy}(z-C_1)dz\right\} = 0,$$

the *fundamental relation*

$$\frac{\psi_1(x-0)+\psi_1(x+0)}{2} - \frac{\psi_1(x_0-0)+\psi_1(x_0+0)}{2}$$

$$= \lim_{y=+0} R\left\{\frac{i}{\pi}\int_{x_0+iy}^{x+iy}\frac{1}{F(z)}dz\right\} \quad (a < x_0,\ x < b). \tag{56}$$

We may change, if necessary, the values of $\psi_1(x)$ at its points of discontinuity, so as to have on (a, b) $\psi_1(x) \equiv \frac{1}{2}(\psi_1(x-0)+\psi_1(x+0))$. Then formula (56) becomes

$$\psi_1(x) - \psi_1(x_0) = \lim_{y=+0} R\left\{\frac{i}{\pi}\int_{x_0+iy}^{x+iy}\frac{1}{F(z)}dz\right\} \quad (a < x_0,\ x < b). \tag{57}$$

If the explicit expression of the function

$$G(z) = \int_a^b \frac{d\psi(u)}{z-u},$$

which, for z not in (a, b), coincides with $F(z)$, is known and is such that $1/G(z)$ is regular analytic inside (a, b), with perhaps a finite number of singularities therein, we can obtain the explicit expression for $\psi_1(x)$, using (57), as follows: we choose x and x_0 sufficiently close to x so that $1/G(x)$ has no singularities on the whole segment (x_0, x). We have then, by Cauchy's Theorem,

$$\frac{i}{\pi}\int_x^{x_0}\frac{dx}{G(x)} + \frac{i}{\pi}\int_0^y\frac{dy}{G(x_0+iy)} + \frac{i}{\pi}\int_{x_0+iy}^{x+iy}\frac{dz}{G(z)} + \frac{i}{\pi}\int_y^0\frac{dy}{G(x+iy)} = 0.$$

Hence

$$R\left\{\frac{i}{\pi}\int_{x_0+iy}^{x+iy}\frac{dz}{G(z)}\right\} = R\left\{\frac{i}{\pi}\int_{x_0}^{x}\frac{dx}{G(x)}\right\}$$

$$- R\left\{\frac{i}{\pi}\int_0^y\frac{dy}{G(x_0+iy)}\right\} + R\left\{\frac{i}{\pi}\int_0^y\frac{dy}{G(x+iy)}\right\}. \tag{58}$$

Now let $y \to +0$. The last two integrals will approach zero, for

$$\left|\frac{i}{\pi}\int_0^y\frac{dy}{G(c+iy)}\right| \leqq \frac{1}{\pi}M_y;\ c = x_0,\ x,\ M = \max\left|\frac{1}{G(x)}\right|,$$

in the rectangle under consideration. Thus, combining (57) and (58):

$$\psi_1(x) - \psi_1(x_0) = R\left\{\frac{i}{\pi}\int_{x_0}^{x} \frac{dx}{G(x)}\right\}. \tag{59}$$

It follows that, if $1/G(x)$ is regular analytic over the whole interval (a, b), with the possible exception of the end points, (59) holds for any interval $(x_0, x) \subset (a, b)$. Then

$$p_1(x) = R\left\{\frac{i}{\pi}\frac{1}{G(x)}\right\} \quad (a < x < b) \quad (d\psi_1(x) = p_1(x)dx). \tag{60}$$

We shall illustrate these considerations with the following examples:

(i) $(a, b) \equiv (-1, 1)$; $p(x) = (1-x^2)^{-1/2}$. Here

$$G(x) = \int_{-1}^{1} \frac{dy}{(1 - y^2)^{1/2}(x - y)} = \pi(x^2 - 1)^{-1/2}.$$

The above considerations are applicable and give

$$p_1(x) = R\left\{\frac{i}{\pi}\frac{(x^2 - 1)^{1/2}}{\pi}\right\} = (1 - x^2)^{1/2}\pi^{-2},$$

which agrees with the results obtained above (page 81).

(ii) $(a, b) \equiv (-1; 1)$; $p(x) = (1+x)^{\alpha-1}(1-x)^{\beta-1}$, α, β positive integers. Here

$$\begin{aligned} G(x) &= \int_{-1}^{1} \frac{(1 + y)^{\alpha-1}(1 - y)^{\beta-1}}{x - y} dy, \\ G(x) &= Q(x) + (1 + x)^{\alpha-1}(1 - x)^{\beta-1} \log [(x + 1)/(x - 1)], \end{aligned} \tag{61}$$

where $Q(x)$ is a polynomial with real coefficients, of degree $\leqq \alpha+\beta-3$, or a constant if $\alpha=\beta=1$, which can be written at once by applying the binomial theory to the integrand in (61). Hence, we can again apply (60), with the result

$$p_1(x) = \frac{(1+x)^{\alpha-1}(1 - x)^{\beta-1}}{\left(Q(x)+(1+x)^{\alpha-1}(1-x)^{\beta-1}\log\frac{1+x}{1-x}\right)^2+(1+x)^{2(\alpha-1)}(1-x)^{2(\beta-1)}\pi^2}. \tag{62}$$

Examples:

$$p_1(x) = (1 + x)^2/\left(-4 - 2x + (1 + x)^2\log\frac{1 + x}{1 - x}\right)^2 + (1 + x)^4\pi^2$$

$$(\alpha = 3, \beta = 1, (a, b) = (-1, 1));$$

$$p_1(x) = \left[\left(\log\frac{1 + x}{1 - x}\right)^2 + \pi^2\right]^{-1} [11], \quad \alpha = \beta = 1;\ (a, b) = (-1, 1). \tag{63}$$

An expression similar to (63) was obtained, though in a different manner and for another purpose, by T. Carleman [11].

(iii) $(a, b)\equiv(-1, 1)$, $p(x)=(1+x)^{\alpha-1}(1-x)^{-\alpha}$ $(1>\alpha>0)$. From (36) and (30) we see that in this case $k_1=0$, so that (42), (60) yield here

$$F(x) = cp(x) \qquad (c = \text{const.}), \tag{64}$$

$$p_1(x) = 1/p(x) \qquad \text{(within a constant factor)}. \tag{65}$$

This agrees with the result on page 82, obtained there in an entirely different manner. If, for example, we take $\alpha=\beta=1/2$, [10], then

$$F(x) = \int_{-1}^{1} \frac{dy}{(1-y^2)^{1/2}(x-y)} = (x^2-1)^{-1/2},$$

and consequently by (65) $p_1(x)=(1+x^2)^{1/2}$ (see page 81).

BIBLIOGRAPHY

[1] Stieltjes, *Recherches sur les fractions continues*, Oeuvres, vol. 2, pp. 398–566.

[2] O. Perron, *Die Lehre von den Kettenbrüchen*, 2d edition, 1929, Chapters VII, VIII, IX.

[3] H. Hamburger, *Ueber eine Erweiterung des Stieltjesschen Momentenproblems*, Mathematische Annalen, vol. 81 (1920), pp. 234–319, vol. 82 (1921), pp. 120–164, 168–187.

[4] J. Shohat and J. Sherman, *On the numerators* · · · , Proceedings of the National Academy of Sciences, vol. 18 (1932), pp. 283–287.

[5] Jacques Chokhate (J. Shohat), *Sur le développement de l'intégrale*

$$\int_a^b \frac{p(y)dy}{x-y} \cdots,$$

Rendiconti di Palermo, vol. 47 (1923), pp. 25–46.

[6] J. Shohat, *On the Stieltjes continued fraction*, American Journal of Mathematics, vol. 54 (1932), pp. 79–84.

[7] Jacques Chokhate (J. Shohat), *Sur une classe étendue de fractions continues* · · · , Comptes Rendus, vol. 191 (1930), p. 988.

[8] Christoffel, *Über die Gaussische Quadratur* · · · , Crelle, vol. 55 (1858), pp. 61–83, where the differential equation for the Legendre case $\alpha=\beta=1$ is obtained in an entirely different manner.

[9] P. Appell et J. Kampé de Fériet, *Fonctions Hypergéométriques et Hypersphériques. Polynomes d'Hermite*, Paris, 1926, p. 338.

[10] C. Possé, *Sur quelques Applications des Fractions Continues Algébriques*, St. Petersbourg, 1886.

[11] T. Carleman, *Sur la résolution de certaines équations intégrales*, Arkiv för Matematik, Astronomi och Fysik, vol. 16 (1922), No. 26.

UNIVERSITY OF PENNSYLVANIA,
PHILADELPHIA, PA.

THREE-DIMENSIONAL MANIFOLDS AND THEIR HEEGAARD DIAGRAMS*

BY

JAMES SINGER

Introduction

One of the outstanding problems in topology today is the classification of n-dimensional manifolds, $n \geqq 3$. Poincaré, the founder of modern analysis situs, devoted several papers to it and allied problems.† Heegaard‡, in a paper concerned primarily with another aspect of the subject, found it convenient to construct a pseudo-normal form for a 3-dimensional manifold, a form which we now call the *Heegaard diagram*. Dehn§ and Veblen|| gave modifications of his construction.

The Heegaard diagram of a 3-dimensional manifold consists of a closed 2-dimensional manifold upon which are drawn a certain number of non-intersecting simple closed curves. Any diagram is an adequate representation of a 3-dimensional manifold in the sense that it completely determines such a manifold, but, unfortunately, a 3-dimensional manifold gives rise to an infinity of diagrams. The problem of classifying manifolds is thus transferred to the problem of classifying diagrams.

Heegaard, in the paper cited above, studied (although not completely) the modifications that can be made on the curves and surface of a diagram which transformed it into another diagram but yet did not change the manifold which it represented. In this paper we extend Heegaard's results and study more completely the relationships between manifolds and their diagrams.

We begin then (Part I) by introducing the notions of a *canonical region*, *canonical surface* and *canonical curve* of a manifold. The *Heegaard diagram* is then constructed from a canonical surface and curves. Then, before proceeding to a discussion of manifolds and their diagrams, we show (Part II) how to read off the usual invariants of a manifold from any one of its representative diagrams.

* Presented to the Society, October 29, 1932; received by the editors June 10, 1932.

† Poincaré, H. For references see his paper in the Rendiconti del Circolo Matematico di Palermo, vol. 18 (1904).

‡ Heegaard, P. A translation into French is given in the Bulletin de la Société Mathématique de France, vol. 44 (1916).

§ Dehn, M., Mathematische Annalen, vol. 69 (1910).

|| Veblen, O., The Cambridge Colloquium, *Analysis Situs*, 2d edition, p. 155.

We then define (Part III) a set of *moves* which operate on the curves and surface of a diagram and transform it into another. Two diagrams are called *equivalent* if one can be obtained from the other by a finite number of these moves. We then prove by a sequence of theorems (Part IV) in which we make use of a specially constructed canonical surface and diagram that any two representative diagrams of a manifold are equivalent. Several other theorems (Part V) lead up to the general theorem of equivalence to the effect that equivalent manifolds (equivalent in the sense of semi-linear analysis situs) arise from and give rise to equivalent Heegaard diagrams (equivalent in our sense) and vice versa.

I wish to thank Professor J. W. Alexander for his many suggestions in the preparation of this paper.

I. Preliminary definitions

1. We will assume that the reader is familiar with the simplex, complex, manifold, incidence matrix, etc., as used in combinatorial analysis situs as, for example, in the book by Veblen, loc. cit. Throughout this paper, simplexes, cells, etc., will be at most 3-dimensional.

2. We derive here several elementary properties of the simplex which we will need later. Let $\bar{\sigma}_3 = A^0A^1A^2A^3$ be a closed 3-simplex, σ_k, σ_{2-k} ($k=0$ or 1) two opposite faces of σ_3 (e.g., $\sigma_k = A^0$, $\sigma_{2-k} = A^1A^2A^3$); let $\bar{\sigma}_3'$ be the derived complex (i.e., a regular subdivision) of $\bar{\sigma}_3$. Every 3-simplex of σ_3' contains a vertex of σ_k or of σ_{2-k}, but not of both. It follows at once that all the 3-simplexes of $\bar{\sigma}_3'$ fall into two groups, R_1 and R_2, where R_1 contains all the simplexes incident with σ_k and R_2 contains all those incident with σ_{2-k}. Moreover, R_1 and R_2 have as boundaries B_1+B and B_2+B, respectively, where B_1 and B_2 consist of 2-simplexes of $\bar{\sigma}_3'$ on the boundary of $\bar{\sigma}_3$, and the common portion B is a 2-cell whose boundary is a circuit on the boundary of $\bar{\sigma}_3$ which "separates" σ_k and σ_{2-k}.

3. **The canonical region, cell and curve.** Given a 3-cell with its boundary sphere S_2, we may decompose S_2 into a sum $E_2' + E_2'' + \bar{B}$ where E_2' and E_2'' are two cells such that $\bar{E}_2'$ and $\bar{E}_2''$ do not meet and B is a spherical band.* We can think of this system as a handle with the E's as its bases.

We will now take a euclidean 3-sphere and attach p handles to it thus obtaining a 3-dimensional region R which we call *canonical*. Explicitly, we have $R = E_3^0 + \sum_{i=1}^{p}(E_3^i + E_2'^{\,i} + E_2''^{\,i})$ where E_3^0 is the spherical region and the elements in the sum represent the handles and their bases. The (point set)-boundary of R will be designated by L.† Incidentally, L is not to be construed

* The bar over a symbol for a set denotes the closure of the set.

† We shall hereafter omit the words *point set*, understanding that whenever we speak of the boundary of a canonical region, we mean the (point set)-boundary.

to lie in euclidean 3-space, for it may very well happen that L is non-orientable.

We recognize in R the following property: if we sever each handle by a cross cut in the form of a 2-cell E_2^i, what is left is a 3-cell. In other words we can think of a canonical region as a region R within which there are p 2-cells E_2^i, with boundaries e^i on the boundary L of R (no two $\bar{E}$'s intersecting) such that $R-\sum E_2^i$ is a 3-cell. The cells E_2^i will be called *canonical cells* and their boundaries *canonical curves*.

4. **The canonical surface and the Heegaard diagram.** A surface L is said to be a *canonical surface* of a manifold M if it satisfies these conditions:

(a) L is a subcomplex of M and is a closed, connected 2-dimensional manifold;

(b) M can be decomposed into R_1+L+R_2, where R_1 and R_2 are canonical regions with the common boundary L.

We note four properties of the canonical surface and regions which follow directly from their definitions:

A. If M' is a subdivision of M, and L', R_1' and R_2' the induced subdivisions of L, R_1 and R_2, then L' is a canonical surface of M' dividing it into the canonical regions R_1' and R_2'. We will also call L' a canonical surface of M.

B. The number of 2-cells that must be removed from R_1 to reduce it to a 3-cell is the same as the number that must be removed from R_2, and each is precisely the maximum number of non-intersecting circuits that can be drawn on L without disconnecting it (Heegaard).

C. R_1 and R_2 are homeomorphic, since they have the same boundary.

D. L and M are both orientable or non-orientable (Heegaard).

5. A problem as yet unsolved is the determination of the minimum genus (or connectivity) of all the canonical surfaces of a given manifold M. This number is clearly a topological invariant. A simpler question is to ask: under what conditions can the genus of a canonical surface of a manifold be lowered or raised? Lemmas 1 and 2 below do not state the most general circumstances but they are sufficient for our needs.

6. In this paragraph we shall use the following notation: Let R be a canonical region, L its boundary; E_3, E_3' and $\mathcal{E}_3$ 3-cells with boundary spheres S_2, S_2' and $\mathfrak{S}_2$. Let $\mathfrak{S}_2$ be separated by a circuit into the two 2-cells $\mathcal{E}_2$ and $\mathcal{E}_2'$, and let $E_3[E_3']$ be separated by two non-intersecting circuits into two 2-cells E_2^1, E_2^2 $[E_2'^1, E_2'^2]$ and a band B $[B']$. Finally, let B $[B']$ be separated into two 2-cells E_2^{*1}, E_2^{*2} $[E_2'^{*1}, E_2'^{*2}]$ by two arcs, each running from one of its bounding circuits to the other.

LEMMA 1a. *If $\bar{\mathcal{E}}_3$ and $R+L$ have in common only $\bar{\mathcal{E}}_2$ on their boundaries, then $R+\mathcal{E}_3+\mathcal{E}_2$ is a canonical region whose boundary is $L+\mathcal{E}_2'-\mathcal{E}_2$.*

LEMMA 2a. *If* $\mathcal{E}_3$ *is a subcomplex of* R *and if* $\mathcal{S}_2$ *and* L *have* $\overline{\mathcal{E}}_2'$ *in common, then* $R-\mathcal{E}_3-\mathcal{E}_2$ *is a canonical region whose boundary is* $L+\mathcal{E}_2-\mathcal{E}_2'$.

These lemmas need no proofs. Obviously, the genus of L is neither raised nor lowered by the operations of the lemmas.

We can, however, change the genus of L by removing or adding a handle, E_3.

LEMMA 1b. *If* E_3 *is a subcomplex of* R *and if* S_2 *and* L *have in common* $\overline{B}$, *then* $R-(E_3+E_2^1+E_2^2)$ *is a canonical region whose boundary is* $L+(E_2^1+E_2^2-B)$.

LEMMA 2b. *If* $\overline{E}_3$ *and* $R+L$ *have in common only* $\overline{E}_2^1$ *and* $\overline{E}_2^2$ *on their boundaries, then* $R+(E_3+E_2^1+E_2^2)$ *is a canonical region whose boundary is* $L+(B-E_2^1-E_2^2)$.

These lemmas, too, need no proof. In the first case, the genus of L is lowered by the removal of a handle, in the second, raised by the addition of a handle.

We can lower the genus of L by attaching to it a 3-cell E_3' along a band and we can raise the genus by removing such a 3-cell, i.e. by "boring" a hole through the region. However, we can add or remove a 3-cell only under certain conditions which are stated in the lemmas below.

LEMMA 1c. *Let* E_3' *and* $R+L$ *have no points in common and* S_2' *and* L *have* $\overline{B}'$ *in common, and let* E_3 *be as in Lemma* 1b; *then, if* S_2 *and* S_2' *have only* $\overline{E}_2^{*1}\equiv\overline{E}_2'^{*1}$ *in common,* $R+E_3'+B'$ *is a canonical region whose boundary is* $L+(E_2'^1+E_2'^2-B')$.

For since, by Lemma 1b, $R-(E_3+E_2^1+E_2^2)$ is a canonical region whose boundary has the closure of the 2-cell $E_2^1+E_2^2+E_2'^{*2}$ in common with the boundary of the 3-cell $E_3+E_3'+E_2{}^{*1}$, it follows from Lemma 1a that

$$\begin{aligned}[R-(E_3+E_2^1+E_2^2)\,]+[E_3+E_3'+E_2^{*1}]&\\ +[E_2^1+E_2^2+E_2'^{*2}]&\equiv R+E_3'+B'\end{aligned}$$

is a canonical region. It is clear that the boundary of $R+E_3'+B'$ is $L+E_2'^1+E_2'^2-B'$.

LEMMA 2c. *If* E_3' *is a subcomplex of* R *such that* S_2' *and* L *have* $\overline{E}_2'^1$ *and* $\overline{E}_2'^2$ *in common, and if there exists a 3-cell* E_3, *subcomplex of* R, *such that* $E_2^{*1}\equiv E_2'^{*1}$ *and* E_2^{*2} *is a 2-cell on* L, *then* $R-E_3'-B'$ *is a canonical region whose boundary is* $L-E_2'^1-E_2'^2+B'$.

For, since $E_3+E_3'+E_2'^{*1}$ is a 3-cell, it follows from Lemma 2a that $R-(E_3+E_3'+E_2'^{*1})-(E_2^1+E_2^2+E_2'^{*2})$ is a canonical region. Hence, by

Lemma 2b, $R-(E_3+E_3'+E_2'^{*1})-(E_2^1+E_2^2+E_2'^{*2})+E_3+E_2^1+E_2^2 \equiv R-E_3'-B'$ is a canonical region, whose boundary, as can be readily seen, is $L-E_2'^1-E_2'^2+B'$.

7. Let the canonical surface L of a manifold M divide it into the two regions R_1 and R_2; let E and F be canonical 2-cells of R_1 and R_2, respectively, e and f, their boundaries. Let E_3 be a 3-cell as in Lemma 1b, where we put $E\equiv E_2^1$ and $R=R_1$. Let A be a 1-cell interior to R_1 with end points on L. Let E_3' be a 3-cell as in Lemma 2c which is a neighborhood of A in R_1. The 2-cells $E_2'^1$ and $E_2'^2$ will be neighborhoods on L of the end points of A. We now have

LEMMA 1. *If the canonical curves e and f meet once and only once, then $L+(E_2^1+E_2^2-B)$ is a canonical surface dividing M into the two canonical regions $R_1-(E_3+E_2^1+E_2^2)$ and $R_2+(E_3+B)$.*

LEMMA 2. *If there exists a 1-cell A' on L such that $A+A'$ bounds a 2-cell of R_1, then $L-(E_2'^1+E_2'^2-B')$ is a canonical surface dividing M into the canonical regions $R_1-(E_3'+B')$ and $R_2+(E_3'+E_2'^1+E_2'^2)$.*

The proofs of the two lemmas follow at once from Lemmas 1 abc, 2 abc. We note that the effect of the first lemma is to remove a handle from R_1 (as in Lemma 1b) and to add a 3-cell to R_2 (as in Lemma 1c) and the effect of the second lemma is to remove a 3-cell from R_1 (as in Lemma 2c) and to add a handle to R_2 (as in Lemma 2b). The genus of the canonical surface is lowered in the first lemma, raised in the second.

8. **Construction of the Heegaard diagram.** The utility of the canonical surface and curves lies in the fact that they give us an adequate representation of the manifold. Indeed, let L be a canonical surface of a manifold M, and let $e^1, \cdots, e^p, f^1, \cdots, f^p$ be two sets of canonical curves, boundaries of canonical sets of 2-cells of R_1 and R_2, respectively. Then, having $L, e^1, \cdots, e^p, f^1, \cdots, f^p$, we can dispense with the rest of M entirely, for any information that can be derived from M can be derived from them. To reconstruct a 3-dimensional manifold, we attach 2-cells to each of the canonical curves, and then add two 3-cells in the obvious way. We thus obtain a manifold N which in general will not be identical to M, but which must always be homeomorphic to it because of the construction.

Of great use in the study of manifolds is the fact that a model of the canonical surface and curves of any manifold can be constructed in ordinary spherical 3-space, S_3. If the canonical surface were orientable, we could immerse it in S_3 immediately; this is impossible if it is non-orientable. To treat both cases at the same time, we adopt one of the normal forms for a 2-dimensional manifold which can be immersed in S_3, i.e., the plane (plus a

point at infinity) from which the interiors of $2p$ circles have been removed.* We call the surface Λ, its $2p$-bounding circles, ϵ_1^i, ϵ_2^i, $i=1, \cdots, p$. For the sake of definiteness later on, let us assume that the circles are of equal radii with centers equally spaced along a straight line.

We now establish a continuous correspondence between the points of L and Λ which is (1, 1) everywhere except that a point on e^i has for image a point on ϵ_1^i and a point on ϵ_2^i. Thus, the image of an e curve will be a pair of the ϵ circles on Λ. If a particular f^i does not meet any of the e curves, then its image on Λ will be a circuit; if it does meet the e's, then its image on Λ will consist of a set of arcs, each joining two points on the circles. We call the circuit or aggregate of arcs corresponding to f^i, ϕ^i.

From what has been said above, it is clear that Λ, ϵ_1^i, ϵ_2^i, and ϕ^i $(i=1, \cdots, p)$ also serve as an adequate representation of the manifold M. We call this representation a *Heegaard diagram* of M.

For some purposes it is convenient to have a Heegaard diagram in which the f's are mapped on the pairs of circles, and the e's become aggregates of arcs. In such cases, we shall use another plane Λ', and introduce notation as needed.

To reconstruct a manifold N from a Heegaard diagram Δ, we first subdivide Λ, if necessary, and then construct a closed 2-dimensional manifold L equivalent to Λ where, however, a single circuit e^i corresponds to the pair of circles ϵ_1^i and ϵ_2^i. We must take care that ϵ_1^i and ϵ_2^i are matched upon e^i with the proper orientations. We then proceed as before, successively adding the 2-cells and finally the two 3-cells.

If a Heegaard diagram is constructed from a manifold, we shall say that the manifold *gives rise* to the diagram and that the diagram *arises from* the manifold; similarly we shall say that a diagram *gives rise* to a manifold and that the manifold *arises from* the diagram when the manifold is constructed from the diagram.

9. Since a Heegaard diagram is an adequate representation of its manifold, all invariants of the latter should be obtainable from the former. We show how to get the Poincaré group, homology characters, and some intersection invariants in the next section.

Since any one manifold can give rise to a great variety of diagrams, we do not seem to be any nearer the solution of the problem of the classification of manifolds. We reserve all such questions for Parts III, IV, and V; at the present we merely note that the form of a Heegaard diagram arising from a manifold is by no means unique.

* See Alexander, J. W., *Normal forms for one- and two-sided surfaces*, Annals of Mathematics, (2), vol. 16, No. 4, June, 1915.

II. The known invariants

1. The Poincaré group. Let Δ be a Heegaard diagram of a manifold M, where Δ consists of Λ, ϵ_1^i, ϵ_2^i and ϕ^i, $i=1, 2, \cdots, p$, all defined as in Part I.

It is well known that there exist, on Λ, $2p$ circuits, a_i and b_i, $i=1, \cdots, p$, all passing through a fixed point O but having no other points in common and such that every other closed curve of Λ is deformable into a sum of the a's and b's. The a's and b's can then be taken as the generators of the Poincaré group of Λ. Moreover, we can always choose the curves in such a manner that a_k is isotopic to ϵ_1^k on Λ and b_k consists of two arcs joining O to congruent points of ϵ_1^k and ϵ_2^k. We can choose a similar base on Λ' (another representation of Δ, in which a pair of circles ϕ_1^i and ϕ_2^i represents a canonical curve f^i), namely c_i and d_i, $i=1, \cdots, p$, where c_k is isotopic to ϕ_1^k on Λ' and d_k consists of two arcs joining O' (image of O) to congruent points of ϕ_1^k and ϕ_2^k.

Since the surfaces Λ and Λ' are representations of the same surface L of M, we can express every curve of one base as a product of the generators of the other base, i.e., $c_k = a_{i_1}^{\epsilon_1} b_{i_2}^{\epsilon_2} a_{i_3}^{\epsilon_3} b_{i_4}^{\epsilon_4} \cdots$, where i_j is one of the integers $1, 2, \cdots, p$ and ϵ_j is 0, 1 or -1. Symbolically we can write

$$c_i = \Pi_i^1 ab, \quad d_i = \Pi_i^2 ab, \tag{1}$$

$$a_i = \Pi_i^3 cd, \quad b_i = \Pi_i^4 cd. \tag{1'}$$

The only identity relation among the generators is

$$a_1 b_1 a_1^{-1} b_1^{-1} a_2 b_2 a_2^{-1} b_2^{-1} \cdots a_p b_p a_p^{-1} b_p^{-1} = 1 \tag{2}$$

in case A is orientable, or

$$a_1 b_1 a_1 b_1^{-1} a_2 b_2 a_2 b_2^{-1} \cdots c_p b_p a_p b_p^{-1} = 1 \tag{2'}$$

in case A is non-orientable.* There are, of course, similar expressions in terms of the c's and d's.

Let us now adjoin to Λ the pairs of congruent 2-cells E_1^i and E_2^i, interiors of the circles ϵ_1^i and ϵ_2^i, where E_1^i and E_2^i correspond to the same canonical 2-cell E^i of R_1. To note what modification the group of Λ undergoes, we shall make use of a theorem concerning the Poincaré group of an arbitrary 2-dimensional complex K_2: If the Poincaré group of K_2 is generated by the elements

$$g_1, g_2, \cdots, g_p$$

with the identity relations

$$r_i = g_{i j_1}^{\epsilon_1} g_{i j_2}^{\epsilon_2} g_{i j_3}^{\epsilon_3} \cdots = 1 \qquad (i = 1, 2, \cdots, q)$$

---

* The identity relation in the non-orientable case is not one of the usual forms; it is, however, equivalent to them in the sense that we can so choose the generators of the Poincaré group that (2′) holds.

where $g_{ij_k}^{e_k}$ is either a generating element or its inverse and we adjoin to K_2 a 2-cell E having no points in common with K_2 but whose boundary is h, where h is a circuit of K_2, then the group of K_2+E is the same as the group of K_2 plus the additional relation

$$r = g_{i_1}^{e_1} g_{i_2}^{e_2} g_{i_3}^{e_3} \cdots = 1,$$

where r represents a curve isotopic to h on K_2.* In particular, if h is isotopic to a generator, say g_1, of the group of K_2, the group of K_2+E can be obtained by replacing g_1 by 1 wherever it occurs in the identity relations of the group of K_2.

It follows at once that the group of $\Lambda+\sum_{i=1}^{p}(E_1^i+E_2^i)$ is generated by the p elements

$$(3) \qquad b_1, b_2, \cdots, b_p$$

with no generating relations, since (2) or (2′) reduces to the unit element identically. The products (1) and (1′) take on new forms which we write symbolically as

$$(4) \qquad c_i = \Pi_i'^1 b, \qquad d_i = \Pi_i'^2 b,$$

$$(4') \qquad 1 = \Pi_i'^3 cd, \qquad b_i = \Pi_i'^4 cd.$$

We now add the p 2-cells F^i to $\Lambda+\sum(E_1^i+E_2^i)\equiv\Lambda+\sum E^i$ obtaining the 2-dimensional complex $L+\sum E^i+\sum F^i$. The products (4) and (4′) take on the new forms

$$(5) \qquad 1 = \Pi_i''^1 b, \qquad d_i = \Pi_i''^2 b,$$

$$(5') \qquad 1 = \Pi_i''^3 d, \qquad b_i = \Pi_i''^4 d.$$

From what has just been said, it is clear that the group of $\Lambda+\sum E^i+\sum F^i$ is generated by the p elements (3) connected by the p relations $\Pi_i''^1\, b=1$ of (5). We obtain an equivalent Poincaré group if we take for basis the p elements

$$(3') \qquad c_1, c_2, \cdots, c_p$$

connected by the p relations $\Pi_i''^3\, d=1$ of (5′).

The addition of the remainder of the manifold M (that is, the two 3-cells) can have no effect on the Poincaré group since a curve homotopic to a point

* If r' is another product of the g's also representing a curve isotopic to h on K_2, then the group obtained by adding r' is equivalent to the group obtained by adding r, since $1=r'=srs^{-1}$, where s is a product of g's.

over M is first deformable over M on to $L+\sum E^i+\sum F^i$ and is then homotopic to a point over $L+\sum E^i+\sum F^i$.

Hence the Poincaré group G of M is given by

$$\begin{array}{lll} & b_1, b_2, \cdots, b_p & \text{generators} \\ b_1^{\alpha_{11}^1} b_2^{\alpha_{12}^1} \cdots b_p^{\alpha_{1p}^1} b_1^{\alpha_{21}^1} \cdots b_p^{\alpha_{2p}^1} \cdots b_p^{\alpha_{k_1 p}^1} & = 1 & \\ b_1^{\alpha_{11}^2} b_2^{\alpha_{12}^2} \cdots b_p^{\alpha_{1p}^2} b_1^{\alpha_{21}^2} \cdots b_p^{\alpha_{2p}^2} \cdots b_p^{\alpha_{k_2 p}^2} & = 1 & \text{identity relations} \\ \cdots\cdots\cdots\cdots\cdots\cdots & & \\ b_1^{\alpha_{11}^p} b_2^{\alpha_{12}^p} \cdots b_p^{\alpha_{1p}^p} b_1^{\alpha_{21}^p} \cdots b_p^{\alpha_{2p}^p} \cdots b_p^{\alpha_{k_p p}^p} & = 1 & \end{array}$$

where the identity relations are the products $\Pi_i^{\prime\prime 1}$ of (5) written out in full. There is an equivalent group in terms of the d's, but we shall not write it out.

2. **The homology characters.** Associated with G^M is the group G^{M_c} which is obtained by making all the elements commutative. G^{M_c} is given by

$$\begin{array}{ll} b_1, b_2, \cdots, b_p & \text{generators} \\ b_1^{\alpha_1^1} b_2^{\alpha_2^1} \cdots b_p^{\alpha_p^1} = 1 & \\ b_1^{\alpha_1^2} b_2^{\alpha_2^2} \cdots b_p^{\alpha_p^2} = 1 & \text{identity relations} \\ \cdots\cdots\cdots\cdots & \\ b_1^{\alpha_1^p} b_2^{\alpha_2^p} \cdots b_p^{\alpha_p^p} = 1 & \end{array}$$

where $\alpha_j^i = \sum_{k=1}^{k_i} \alpha_{kj}^i$.

Every invariant of G^{M_c} is an invariant of M, the manifold giving rise to the Heegaard diagram Δ. In particular, G^{M_c} yields the 1-dimensional Betti and torsion numbers.

If M is orientable, its Betti numbers are 1, $p-\rho$, $p-\rho$, 1; if M is non-orientable, its Betti numbers are 1, $p-\rho$, $p-\rho-1$, 0; where ρ is the rank of the matrix $\|\alpha_j^i\|$. If M is orientable, it has a set of 1-dimensional coefficients of torsion given by the invariant factors of $\|\alpha_j^i\|$; if M is non-orientable, it has in addition to these a 2-dimensional coefficient of torsion equal to two.

3. **The intersection numbers.** We now pass to the intersection invariants of the manifold M. There are two types of such invariants, the Kronecker indices of the 1- and 2-cycles and the intersection 1-cycles of 2-cycles with one another. To find these invariants we need to construct a homology base for the 1- and 2-cycles.

Let us consider the boundary relations

$$F^i \to \phi^i \qquad \text{(on } M\text{)} \tag{6}$$

and the homologies

$$\phi^i \sim \sum_j \alpha_j{}^i b^j \qquad \text{(on } L\text{)}, \tag{7}$$

where the matrix $\|\alpha_j{}^i\|$ is the same as the matrix of exponents in the identity relations of G^{M_c}. It is well known that there exist unimodular transformations $F^i = \sum_j \mu_j^i F'^j$, $\phi^i = \sum_j \mu_j^i \phi'^j$ and $b^i = \sum_j \nu_j^i b'^j$ such that the bounding relations (6) and the homologies (7) become

$$F'^i \rightarrow \phi'^i \qquad \text{(on } M\text{)} \tag{8}$$

and

$$\phi'^i \sim \sum_j \alpha_j'{}^i b'^j \qquad \text{(on } L\text{)} \tag{9}$$

respectively, where $\|\alpha_j'{}^i\|$ is a diagonal matrix. The first $\rho-\tau$ terms of the main diagonal of $\|\alpha_j'{}^i\|$ are equal to 1, the next τ terms are the invariant factors (1-dimensional coefficients of torsion) and all the other terms are zero.

Since every 1-cycle of M is homologous to a linear combination of the b's, and hence to the b''s, and since the 1-dimensional Betti number of M is $p-\rho$, it follows that the curves

$$b'^i \qquad (i = \rho+1, \rho+2, \cdots, p) \tag{10}$$

form a basis for the non-bounding 1-cycles of M in this sense: if b is any 1-cycle of M, then

$$b \sim \sum_{i=\rho+1}^{p} a_i b'^i + \sum_{i=\rho-\tau+1}^{\rho} \bar{a}_i b'^i.$$

But since some multiple of b'^i $(\rho-\tau+1 \leqq i \leqq \rho)$ bounds, it will meet any 2-cycle of M zero times algebraically. Hence as far as the intersection of b with a 2-cycle is concerned, only the terms of the first sum in b need to be considered.

It can be shown that the 2-dimensional complexes

$$F'^i \qquad (i = \rho+1, \rho+2, \cdots, p) \tag{11}$$

can serve as a base for the non-bounding 2-cycles as far as intersections are concerned. To be sure, none of the F's is a cycle, nor is any combination of them a cycle since every sum $\sum \lambda_i F'^i$ has for boundary $\sum \lambda_i \phi'^i$, but it can be shown that if the complexes F_0^i, $i = \rho+1, \cdots, p$, do form a base for the non-bounding 2-cycles of M in the same sense as the b^i's just above, then the intersection of b^i and F_0^j on M is the same as the intersection of b^i and the

boundary of F^i on L. Hence the intersection matrix of the non-bounding 1- and 2-cycles of M is given by

$$(12) \qquad (b'^i \cdot F'^j) \qquad (i, j = \rho + 1, \cdots, p)$$

where the element in the ith row and jth column is the Kronecker index of b'^i and ϕ'^j on Λ. In case M is orientable, the matrix (12) is unimodular and can be transformed into a diagonal matrix by further unimodular transformations of the b''s and F''s.

The 2-complexes (11) also yield the intersections of the non-bounding 2-cycles among themselves. Indeed, we can compute from them a cubic matrix

$$(13) \qquad \|\sigma_{ijk}\| \qquad (i, j = \rho + 1, \rho + 2, \cdots, p;\ k = 1, 2, \cdots, p),$$

where we can show that $\sum_k \sigma_{ijk} b^k$ is a cycle homologous to the intersection cycle of two non-bounding 2-cycles of M, as defined, for example, by Lefschetz in his Colloquium Lectures, Chapter IV. The σ's are precisely the same as those defined by Alexander in Proceedings of the National Academy of Sciences, 1924, p. 99.

They are obtained in this fashion. If $i = j$, then $\sigma_{ijk} = 0$ for all k's. If $i \neq j$, we proceed as follows.

We consider a particular pair of (11), say F'^1 and F'^2, and a particular canonical circle, say ϵ_1^3. Since F'^1 can be considered a two-cycle, its boundary on Λ, namely ϕ'^1, meets ϵ_1^3 zero times algebraically, if we count each intersection with its proper orientation and multiplicity. Now imagine that the curve ϵ_1^3 is shrunk to its center A. The point A is met zero times algebraically by the segments of ϕ'^1 abutting on it.

Similarly, the point A will be met zero times, algebraically, by the segments of ϕ'^2 incident with it.

If no segment is common to ϕ'^1 and ϕ'^2, we define the number σ_{123} as the Kronecker index of ϕ'^1 and ϕ'^2 at the point A, where we take into account the multiplicity of the various branches.

Suppose now that the segment a belongs to ϕ'^1 with the multiplicity m_1 and to ϕ'^2 with the multiplicity m_2. Replace a by two distinct segments a^1 and a^2 lying very close to it, and let a^1 belong to ϕ'^1 with the multiplicity m_1 and a^2 to ϕ'^2 with the multiplicity m_2. If we do this for all segments common to ϕ'^1 and ϕ'^2 we are led to the former case and can define σ_{123} as there. However, it is apparent that σ_{123} as so defined is not unique, for the Kronecker index will depend on the method of replacing a by a^1 and a^2, that is, on whether we take a^1 to the right or left of a^2. We therefore define as the σ_{123}

the number of smallest numerical value so obtained. As a matter of fact, the difference between any two values represents a cycle which is homologous to zero; we choose the smallest numerical value purely for convenience.

We can now define the σ's for all i, j, and k in the same way. It is clear that we need not consider the canonical circles ϵ_2^i, since the intersections are the same as on ϵ_1^i.

III. Definitions of the moves and equivalent Heegaard diagrams

1. Two Heegaard diagrams Δ and Δ' with canonical curves $\epsilon^1, \cdots, \epsilon^p, \phi^1, \cdots, \phi^p$ and $\epsilon'^1, \cdots, \epsilon'^q, \phi'^1, \cdots, \phi'^q$, respectively, are said to be *identical* if

(a) $p=q$, and

(b) Δ and Δ' are mapped on the same euclidean plane Λ in such a way that each pair of circles $\epsilon_1'^k$ and $\epsilon_2'^k$ coincides in position and orientation with a pair $\epsilon_1{}^i$ and $\epsilon_2{}^i$ and each ϕ'^k is isotopic to a ϕ^i on Λ.

2. We now define a set of transformations, called *moves*, which operate on and modify the curves and surface of a Heegaard diagram. The moves will fall into three classes or types; the first will modify the canonical curves, the second will modify the plane Λ by rotating a portion of it through a multiple of π, the third will modify the canonical curves and the plane Λ by the addition (or subtraction) of canonical curves.

Type I. A. A move of Type IA simply changes the orientation of a given canonical curve.

Let g be a 1-sphere on Λ not meeting any of the circles (images of the canonical e curves) and containing in its interior at least one circle, say ϵ_1^k, and at most two such circles. In the latter case, the two circles must be adjacent but not the two circles of a pair, i.e., images of the same e. Sever Λ along g and call the inner and outer lips of the cut g and g', respectively. Remove the interior of g from Λ and attach it once more, this time, however, by matching corresponding points of ϵ_1^k and ϵ_2^k. We have lost the canonical pair ϵ_1^k and ϵ_2^k and have gained a new pair of circuits, g and g'. The two cases give us the moves of Types IB and IC, i.e.,

B. g contains only one circle,

C. g contains two circles.

A move of Type IA, B or C is its own inverse. The effect of a move of Type IB is to replace a canonical curve by one isotopic to it. The effect of a move of Type IC is to replace a canonical curve by the "sum" of it and another canonical curve.

D. A move of Type ID is the replacement of the image on Λ of a canonical f curve, say ϕ^i, by the sum of the images of f^i and some f^j, i.e., by $\phi^i+\phi^j$.

The move is effected by deforming ϕ^i on Λ until it has a 1-simplex in common with ϕ^j. Then by removing this 1-simplex we obtain a circuit $\phi'^i=\phi^i+\phi^j$, which after a slight deformation will have no points in common with ϕ^j. The replacement of ϕ^i by ϕ'^i is the move of Type ID.

When a diagram has been modified by a move of Type IB or C, its circles will no longer be in standard form, i.e., of equal radii and equally spaced along a straight line. However, an isotopic deformation of Λ in S_3 will bring them into standard form. We shall always suppose this done when we operate on Λ by a move of Type IB or C.

Type II. Let g be a circuit on Λ having no points in common with the canonical set $\{e\}$ and let g' be its position after a small isotopic deformation such that g and g' have no points in common, i.e., they have the appearance of two concentric circles, where g' is interior to g, say. Let B be the band-shaped region bounded by g and g'. Now rotate g' and that part of Λ interior to it through a positive angle of π or 2π, keeping g and that part of Λ exterior to it fixed. We note particularly that if one of a pair of canonical circles lies within g' and the other without, only one of them is rotated. The result is a distortion in the band B; however, if B is suitably subdivided into simplexes of fine enough mesh, its structure will remain unaltered.

However, we do not wish to employ this type of transformation in its most general form but only in two certain cases when the region interior to g' contains either one or two canonical curves.

A. Rotation through a positive angle π when the 1-sphere g' contains two consecutive circles.

B. Rotation through a positive angle 2π when the 1-sphere g' contains only a single circle.

Inverses of moves of Types IIA and B will be rotations through negative angles.

Type III. A move of this type adds a pair of canonical circles and a new canonical curve to the plane Λ. It is effected as follows: Let ϵ_1^{p+1} and ϵ_2^{p+1} be a pair of circuits on Λ each bounding a 2-cell, and let ϕ^{p+1} be an arc joining the point P_1 of ϵ_1^{p+1} and the point P_2 of ϵ_2^{p+1}. The arc ϕ^{p+1} must have no other points in common with ϵ_1^{p+1} or ϵ_2^{p+1} and no points in common with the two 2-cells bounded by ϵ_1^{p+1} and ϵ_2^{p+1}. Also ϵ_1^{p+1}, ϵ_2^{p+1} and ϕ^{p+1} must not meet any of the other canonical curves or arcs. Then if we remove the interiors of ϵ_1^{p+1} and ϵ_2^{p+1} from Λ and identify the points of ϵ_1^{p+1} and ϵ_2^{p+1} so that P_1 and P_2 are matched, we can add the pair ϵ_1^{p+1} and ϵ_2^{p+1} to our canonical circles and ϕ^{p+1} to our canonical arcs. Its inverse, the removal of a pair of canonical curves ϵ^{p+1} and ϕ^{p+1} under the conditions just described, will be denoted by III'.

The surface Λ seems to play a preferred rôle in the description and definition of these moves. However, all our moves are also applicable to the surface Λ', where the ϕ curves are represented by pairs of circles. It is not necessary to note the effect on Λ' of one of the moves operating on Λ; we only note that when we modify Λ by a move of Type IC, Λ' is modified by a move of Type ID, and vice versa.

3. Let now Δ be a Heegaard diagram with the two canonical sets of curves $\{\epsilon\}$ and $\{\phi\}$. We state six lemmas whose proofs follow immediately from the definitions of the moves.

LEMMA 3. *If g is a circuit on Λ not meeting any of the canonical circles and containing in its interior ϵ_1^k and any number of other circles but not ϵ_2^k, then the result of severing Λ along g and reattaching the piece so removed to Λ along ϵ_1^k and ϵ_2^k may be obtained by successively applying moves of Type* IC.

LEMMA 4. *If the set of curves $\{\epsilon'\}$ is derived from the canonical set $\{\epsilon\}$ by a finite number of moves of Type* I, *then the set $\{\epsilon'\}$ is canonical.*

LEMMA 5. *If the canonical set $\{\epsilon'\}$ is derived from the canonical set $\{\epsilon\}$ by a finite number of moves of Type* I, *then $\{\epsilon\}$ can be derived from $\{\epsilon'\}$ by the same number of moves.*

LEMMA 6. *If the surface Λ and the canonical sets of curves $\{\epsilon\}$ and $\{\phi\}$ of a Heegaard diagram Δ are transformed into a surface Λ' with the sets of curves $\{\epsilon'\}$ and $\{\phi'\}$ by a finite number of moves, then Λ', $\{\epsilon'\}$ and $\{\phi'\}$ form a Heegaard diagram.*

LEMMA 7. *If the Heegaard diagram Δ' is derived from the Heegaard diagram Δ by a finite number of moves, then Δ can be derived from Δ' also by a finite number of moves.*

LEMMA 8. *If the Heegaard diagram Δ gives rise to the manifolds N and N', then N and N' are homeomorphic.*

COROLLARY. *If the manifold M gives rise to the diagram Δ, which in turn gives rise to the manifold N, then M and N are homeomorphic.*

The proofs of this lemma and corollary follow at once from the construction of N and N' as outlined in §8, Part I.

4. We are now in a position to give a formal definition of equivalence: two Heegaard diagrams are said to be *equivalent* if one can be transformed into a diagram identical to the other by a finite number of moves of Types I, II and III.

5. We prove the following theorem:

THEOREM 1. *If the equivalent Heegaard diagrams Δ and Δ' give rise to the manifolds N and N', then N and N' are homeomorphic.*

Since Δ and Δ' are equivalent, Δ can be obtained from Δ' by a finite sequence of moves of Types I, II, and III which transform Δ' successively into the diagrams $\Delta^1, \Delta^2, \cdots, \Delta^m \equiv \Delta$. It is therefore necessary and sufficient to prove that at each stage the transformed diagram Δ^{k+1} gives rise to a manifold N^{k+1} homeomorphic to the one, N^k, arising from the preceding diagram Δ^k.

This is obviously true when we modify Δ^k by a move of Type IA, IB, or a move of Type II, hence we need only prove that the theorem holds when we employ a move of Type IC, ID, or a move of Type III.

Let, then, Δ^k be transformed into Δ^{k+1} by a move of Type IC and let Δ^k and Δ^{k+1} give rise to the manifolds N^k and N^{k+1}. We can suppose that the move replaces the canonical curve ϵ_k^1 of Δ^k by the canonical curve $\epsilon_{k+1}^1 = \epsilon_k^1 + \epsilon_k^2$. Let us retain the circuit ϵ_k^1 on the diagram Δ^{k+1} and slightly deform it so that it does not meet any of the canonical curves of Δ^{k+1}. We can then find in the manifold N^{k+1} a 2-cell E_k^1 lying in the region containing the canonical 2-cells E_{k+1}^i and whose boundary is ϵ_k^1. The 2-cell E_k^1 will divide this region, a 3-cell, into two parts. If we unite these two parts along E_{k+1}^1 and consider E_k^1 as part of the boundary, we have again a 3-cell. We have not changed N^{k+1} at all, but obviously from this point of view it may be considered a manifold arising from Δ^k; hence by Lemma 8, N^{k+1} is homeomorphic to N^k.

A similar argument holds when Δ^k is transformed by a move of Type ID.

Suppose now that Δ^k is transformed into Δ^{k+1} by a move of Type III which adds the canonical curves ϵ^{p+1} and ϕ^{p+1} to the canonical curves of Δ^k. In the manifold N^{k+1} the curves ϵ^{p+1} and ϕ^{p+1} are such that Lemma 1 is applicable where the 2-cell E^{p+1} plays the rôle of E_2^1 of the lemma. If we make the necessary modifications, we transform the canonical surface and regions of N^{k+1} into new ones. But with these latter canonical surface and regions N^{k+1} may be considered as a manifold arising from Δ^k; hence, once more N^k and N^{k+1} are homeomorphic and the theorem is proved.

6. If, then, the Heegaard diagrams Δ and Δ' are equivalent, i.e., if one can be transformed into the other by means of the moves, then any two manifolds N and N' that arise from them are homeomorphic. To give greater justification to the definition and notion of equivalence, we must prove conversely that if two manifolds M and M' are homeomorphic and they give rise to the Heegaard diagrams Δ and Δ' then the two diagrams are equivalent. In other words, we must prove that the moves are indeed sufficient to transform Δ into Δ'.

IV. Heegaard diagrams arising from a manifold

1. Variations in a Heegaard diagram may arise in

(a) the choice of the canonical surface,

(b) the choice of the canonical 2-cells of one of the canonical regions,

(c) the choice of the canonical 2-cells of the other canonical region,

(d) the method of immersion in S_3, i.e., the method of mapping L on Λ.

In this section we study the relationship between any two diagrams arising from a manifold and prove (Theorems 2–8) that any two such diagrams are equivalent.

We shall use consistently the following notation: M, M', N, etc., shall denote a 3-dimensional manifold; L, a canonical surface, R_1 and R_2 the two canonical regions. Canonical sets of 2-cells of R_1 shall be denoted by $\{E\}$, $\{E'\}$, etc., of R_2 by $\{F\}$, $\{F'\}$, etc. The canonical sets of curves (boundaries of the canonical 2-cells) will be denoted by $\{e\}$, $\{f\}$, etc. We shall denote a Heegaard diagram by Δ, Δ', etc., and we shall use the notation $\Delta\equiv(\Lambda, \epsilon, \phi)$ to signify that the Heegaard diagram Δ consists of the plane Λ (image of a canonical surface L), the pairs of circles, ϵ_1^i, ϵ_2^i (images of the canonical e curves), and the arcs ϕ^i (images of the canonical f curves).

2. In Theorem 2 below we prove that two diagrams arising from the same canonical surface and curves are equivalent, i.e. two methods of immersing L, $\{e\}$ and $\{f\}$ in S_3 (see (d) of §1) yield equivalent diagrams.

Theorem 2. *Let L be a canonical surface of a manifold M, $\{e\}$ and $\{f\}$ canonical sets of curves of the two regions; then if L, $\{e\}$ and $\{f\}$ give rise to the Heegaard diagrams $\Delta\equiv(\Lambda, \epsilon, \phi)$ and $\Delta'\equiv(\Lambda', \epsilon', \phi')$, Δ and Δ' are equivalent.*

In this theorem Λ and Λ', ϵ^i and ϵ'^i, ϕ^i and ϕ'^i correspond to the same L, e^i and f^i, respectively, of M. Let us superimpose the two planes Λ and Λ'; we can assume, without any loss of generality, that the canonical circles coincide. We choose our notation so that ϵ_1^k, ϵ_2^k and $\epsilon_1'^k$, $\epsilon_2'^k$ (ϕ^k and ϕ'^k) are images of the same e^k (f^k) of L.

The circles ϵ_1^k, ϵ_2^k and $\epsilon_1'^k$, $\epsilon_2'^k$ will not coincide, in general, but we shall show how to modify Λ' by means of our moves so that the two images of each e do coincide and the two images of each f are isotopic.

We regard the two planes Λ and Λ' as infinitely close but distinct, as, for example, two sheets of a Riemann surface. There will be no misunderstanding then when we speak of the intersection of a curve on Λ with a curve on Λ', and at the same time insist that when we modify Λ' by a move, no change occurs on Λ.

Let us denote by λ^i, $i=1, 2, \cdots, 2p-1$, the $2p-1$ segments of the straight line in Λ through the centers of the canonical circles. The segment

λ^k then joins the kth and the $(k+1)$st circles. By subdivision of Λ we can make these segments 1-cells of Λ. Let λ'^i, $i=1, 2, \cdots, 2p-1$, be the images of λ^i in Λ'. Then if λ^1 joins ϵ_1^5 to ϵ_2^2, λ'^1 joins $\epsilon_1'^5$ to $\epsilon_2'^2$.

Since $\Lambda-\sum_{i=1}^{p}(\epsilon_1^i+\epsilon_2^i)-\sum_{j=1}^{2p-1}\lambda^j$ is a 2-cell, to prove Δ is equivalent to Δ' it is sufficient to prove that we can modify Λ' by our moves so that ϵ_1^k and ϵ_2^k coincide with $\epsilon_1'^k$ and $\epsilon_2'^k$, respectively, and λ^k is isotopic to λ'^k. But this is quite obviously possible by means of moves IIA and IIB. Indeed, let ϵ_1^1 be the first circle on Λ (reading from left to right, say). The 1-cell λ^1 joins ϵ_1^1 to ϵ^a, say. By a finite number of moves of Type IIA (acting on $\epsilon_1'^1$ of Λ') we can make $\epsilon_1'^1$ coincide with ϵ_1^1. Then by another finite sequence of moves we can make ϵ'^a coincide with ϵ^a. Moreover, we can choose this sequence of moves in such a fashion that λ'^1 loses all its intersections with all λ^i, $i\neq 1$. Then by a finite number of moves of Type IIB acting on ϵ'^a, we can make λ'^1 lose all its intersections with λ^1. The 1-cells λ^1 and λ'^1 are now isotopic. By continuing this process, we can make each λ^k isotopic to its corresponding λ'^k and each ϵ_1^k (ϵ_2^k) coincident with $\epsilon_1'^k$ $(\epsilon_2'^k)$. The theorem is therefore proved.

Corollary 1. *Let $\Delta\equiv(\Lambda, \epsilon, \phi)$ be a given Heegaard diagram, and let $\bar{\Delta}\equiv(\bar{\Delta}, \bar{\phi}, \bar{\epsilon})$ be a diagram obtained from Δ by mapping the arcs ϕ^i of Λ as pairs of circles on $\bar{\Lambda}$ and the pairs of circles ϵ_1^i and ϵ_2^i of Λ on arcs $\bar{\phi}^i$ of $\bar{\Lambda}$. Then if $\bar{\Delta}'\equiv(\bar{\Lambda}', \bar{\phi}', \bar{\epsilon}')$ is another diagram similarly obtained, $\bar{\Delta}$ and $\bar{\Delta}'$ are equivalent.**

Corollary 2. *Let the diagram Δ be equivalent to the diagram Δ'; then if the diagrams $\bar{\Delta}$ and $\bar{\Delta}'$ are obtained from Δ and Δ', respectively, as above, they are equivalent.*

The proofs of these corollaries follow directly from Theorem 2. We only need to point out that a move of Type IC acting on Δ is equivalent to a move of Type ID acting on $\bar{\Delta}$.

3. Theorem 3 below states that two Heegaard diagrams arising from two choices of the canonical 2-cells of R_1 (see (b) §1) are equivalent.

Theorem 3. *If $\{e\}$ and $\{e'\}$ are two sets of canonical curves for the same region R_1 of M, then any diagram $\Delta\equiv(\Lambda, \epsilon, \phi)$ arising from L, $\{e\}$ and $\{f\}$ is equivalent to any diagram $\Delta'\equiv(\Lambda', \epsilon', \phi')$ arising from L, $\{e\}$ and $\{f\}$.*

In this theorem, Λ and Λ', ϕ^i and ϕ'^i correspond to the same L and f^i, respectively, of M, but ϵ corresponds to an unprimed e whereas ϵ' corresponds to a primed e.

Let ϵ''^i be the image of e^i on Λ'. We will prove the theorem by showing how to modify Λ' so that the ϵ''s are replaced, one by one, by the ϵ'''s. The

* From here on, $\bar{\Delta}$, etc., will indicate not the closure of Δ, etc., but a new Δ, etc.

diagram Δ' will be then transformed into the diagram Δ''. By definition, Δ' and Δ'' are equivalent, by Theorem 2, Δ and Δ'' are equivalent, hence Δ and Δ' are equivalent. The process by which we modify Λ' will be broken up into three cases.

Case 1. We suppose first that no ϵ'' meets an ϵ_1' or an ϵ_2' and that no ϵ'' is interior to another.

The ϵ'' curves will then be circuits on Λ' not meeting any of the circles, $\epsilon_1'^i$ or $\epsilon_2'^i$. Each ϵ'' must contain at least one ϵ' circle; for any i, either $\epsilon_1'^i$ or $\epsilon_2'^i$ or both are contained in an ϵ''; and there is at least one ϵ' circle not contained in any ϵ''.

Let $\epsilon_1'^k$ be a circle not contained in any ϵ''. Its partner $\epsilon_2'^k$ must be contained in at least one ϵ''. Several subcases now arise.

Suppose first only one ϵ'' circle, say ϵ''^a, contains $\epsilon_2'^k$ and that ϵ''^a contains no other ϵ'. Then by a move of Type IB, in which we sever Λ' along ϵ''^a and heal it up again by matching $\epsilon_1'^k$ and $\epsilon_2'^k$, we transform one of the ϵ''s into one of the ϵ'''s.

Suppose secondly that only one ϵ'' circle, say ϵ''^b, contains $\epsilon_2'^k$, but that ϵ''^b contains other ϵ' circles. If ϵ''^b contains only one other such circle, say $\epsilon_1'^m$, then by a move of Type IC by which we sever Λ' along ϵ''^b and heal it up again by matching $\epsilon_1'^k$ and $\epsilon_2'^k$, we again transform one of the ϵ''s into one of the ϵ'''s. The circle ϵ'^m is now not contained in any ϵ'' curve.

If ϵ''^b contains more than one circle besides $\epsilon_2'^k$, we modify Λ' by a move of Type IC by which we sever Λ' along a circuit not meeting ϵ''^b and containing $\epsilon_2'^k$ and one other circle and then patch it up by matching $\epsilon_1'^k$ and $\epsilon_2'^k$. We thus obtain a new configuration in which there are one fewer circles interior to ϵ''^b. One of the new canonical circles is interior and the other is exterior to ϵ''^b. By repeating this process we can remove the circles interior to ϵ''^b one by one until only two remain, and then we are back to the preceding case.

By a finite number of steps we can modify Λ' so that all the ϵ''s become ϵ'''s. Hence, Δ' is transformed into a diagram Δ'' representing L, $\{e\}$ and $\{f\}$, and is therefore equivalent to Δ by Theorem 2, as we wished to prove.

Case 2. Let no ϵ'' meet an ϵ' circle, but suppose that some ϵ'''s are interior to others.

Suppose first that the ϵ'''s form "nests" of circuits, i.e. if ϵ''^a contains ϵ''^b and ϵ''^c, then either ϵ''^b contains ϵ''^c or ϵ''^c contains ϵ''^b. By considerations entirely analogous to those of Case 1 we can modify Λ' so that the innermost ϵ'' of a nest is changed into an ϵ' and hence the nest has one fewer ϵ'' curves. After a finite number of steps we are led back to Case 1.

Now suppose that the ϵ'''s are contained in other ϵ'''s in a general fashion.

Several cases arise, but it is not necessary to go into details. We can always transform Λ' so that the ϵ'''s are transformed into nests of curves and we are back to the former case.

Case 3. We suppose that the ϵ'''s do meet the ϵ''s.

Let us examine, first of all, the nature of the intersection of an unprimed canonical 2-cell, say E^j, with a primed canonical 2-cell, say E'^k, in the manifold M. The intersection will consist of a number of 1-cells and circuits. By a slight deformation of E'^k and its boundary e'^k we can arrange so that e'^k meets e^j, the boundary of E^j, only in a finite number of points, such that the end points of any 1-cell common to E^j and E'^k are distinct.

Keep j fixed and let k run from 1 to p. We obtain on e^j a certain number of points, grouped into pairs, where each pair is the boundary of a 1-cell common to E^j and some E' and no pair "separates" another on e^j. Hence there is at least one pair, say P_1 and P_2, such that one of the two arcs into which P_1 and P_2 divide e^j contains no other intersection point. Let us call this arc a. The points P_1 and P_2 are on some e', say e'^k.

We now return to the Heegaard diagram Δ' on Λ'. The canonical curve e'^k is represented by the pair of circles $\epsilon_1'^k$ and $\epsilon_2'^k$; each circle has on it the images of P_1 and P_2, which we continue to call P_1 and P_2, and a is now mapped on a 1-cell α of Λ' joining P_1 and P_2 of $\epsilon_2'^k$, say. The arc α meets no other circle.

We now show that we can always modify Λ' so as to lose the two intersections P_1 and P_2. The circle $\epsilon_2'^k$ and the arc α divide Λ' into two regions. Let A be the one which does not contain $\epsilon_1'^k$. Several cases arise. Suppose first that A contains no canonical circle at all. Then, obviously, α can be deformed so that the intersections P_1 and P_2 are lost.

Suppose next that A contains only one canonical circle. Let g be a circuit in $\Lambda'-A$ which lies very close to α and that part of $\epsilon_2'^k$ bounding $\Lambda'-A$. Then if we modify Λ' by a move of Type IC in which we sever Λ' along g and patch it up again by matching $\epsilon_1'^k$ and $\epsilon_2'^k$, we find that the two intersections P_1 and P_2 have disappeared.

Suppose, finally, that A contains several canonical circles. Choose g as before and again sever Λ' along it and patch it up again along corresponding points of $\epsilon_1'^k$ and $\epsilon_2'^k$. Again, we lose the two intersections, P_1 and P_2, and, by Lemma 3, this operation is the product of moves of Type IC. Hence since we can remove all the intersections two by two this case is reduced to the former and the theorem is completely proved.

4. Theorem 4 states that two Heegaard diagrams arising from two choices of the canonical 2-cells of R_2 (see (c), §1) are equivalent. In the previous theorem, the circles on Λ and Λ' were images of different e's; in this theorem, the ϕ's are images of different f's.

THEOREM 4. *If* $\{f\}$ *and* $\{f'\}$ *are two sets of canonical curves for the same region* R_2 *of* M, *then any diagram* $\Delta \equiv (\Lambda, \epsilon, \phi)$ *arising from* L, $\{e\}$ *and* $\{f\}$ *is equivalent to any diagram* $\Delta' \equiv (\Lambda', \epsilon', \phi')$ *arising from* L, $\{e\}$ *and* $\{f'\}$.

This theorem follows at once from Corollaries 1 and 2 of Theorem 2 and Theorem 3 by mapping Δ and Δ' on $\bar{\Delta}$ and $\bar{\Delta}'$ as in Corollary 1.

5. We have proved, Theorems 2, 3, and 4, that any two Heegaard diagrams arising from a manifold are equivalent, provided that we chose the same canonical surface in each case. We have left to prove that any two diagrams whatsoever arising from a manifold are equivalent (see (a), §1). To prove this we make use of a special canonical surface whose construction is given below.

6. **Construction of the special canonical surface.** Let M be a manifold which we assume simplicial, G the linear graph consisting of all the 0- and 1-simplexes of M, H a subcomplex of G. Let M' be the first derived complex of M, and M'' the first derived complex of M'; and let G', G'', H', H'' be the induced subdivisions on G and H, respectively. It may happen that the boundary† of the M''-neighborhood of H is a canonical surface, in which case we call it a *special canonical surface* and any diagram arising from it a *special Heegaard diagram.* In particular, if H is G itself, we prove

LEMMA 9. *The boundary,* L, *of the* M''*-neighborhood of* G *is a special canonical surface.*

Let M^* be the dual of M, G^* the linear graph consisting of all the 0- and 1-cells of M^*, and $G^{*\prime}$ the subdivision of G^* induced by M'. Further, let R_1 and R_2 be the M''-neighborhoods of G and G^*, respectively. We prove that L is a canonical surface by showing

(a) that every 3-simplex of M'' belongs either to R_1 or R_2;

(b) L is the common boundary of R_1 and R_2, and is a closed and connected 2-dimensional manifold;

(c) R_1 and R_2 are canonical regions.

It then follows from our definitions (Part I, §4) that L is a canonical surface dividing M into the canonical regions R_1 and R_2.

(a) Let A_0^i be the vertices of M; B_1^j, B_2^k, B_3^l the vertices of M' on the 1-, 2- and 3-cells of M, respectively, and C_1^r, C_2^s, C_3^t the vertices of M'' on the 1-, 2- and 3-cells of M', respectively. Every 3-simplex of M' is of the form $A_0B_1B_2B_3$, every 3-simplex of M'' is of the form $A_0C_1C_2C_3$ or $B_aC_1C_2C_3$, where $a=1, 2$, or 3. Hence every 3-simplex of M'' is incident with an A_0 or a B_1, i.e. a vertex of G' or else with a B_2 or a B_3, i.e. a vertex of $G^{*\prime}$. That is, every 3-simplex of M'' belongs to R_1 or to R_2.

† See second footnote on page 89.

(b) In any 3-simplex, σ_3^i, of M', the component 3-simplexes of R_1 and R_2 are grouped as the simplexes of §2 of Part I, hence in any 3-simplex of M', the boundaries of R_1 and R_2 have in common a 2-cell which we call E_2^i. The boundary of E_2^i is a circuit which may be thought of as composed of four 1-cells, each on one of the faces of σ_3^i. Let us call the 1-cell on the 2-simplex σ_2^i, E_1^i. It is quite clear that the incidence matrix of the 2- and 3-simplexes of M' is the same as the incidence matrix of all the 1- and 2-cells E_1^i and E_2^i. It follows at once from the fact that M is a manifold that $L=\sum E_2^i$ (the sum being taken over all i's) is a closed, connected 2-dimensional cellular manifold, from which it can be deduced that, upon subdivision, L will become a closed, connected 2-dimensional simplicial manifold. Furthermore, it follows that L is orientable or non-orientable according as M is orientable or non-orientable, and conversely.

(c) Since the M''-neighborhood of a 1-simplex of M is a 3-cell, the M''-neighborhood of any tree of 1-simplexes of M is also a 3-cell. Let the linear graph G contain p independent circuits; then the removal of p properly chosen 1-simplexes, say $\sigma_1^1, \sigma_1^2, \cdots, \sigma_1^p$, from G will reduce it to a tree T. Let B_1^i be the vertex of M' on σ_1^i $(i=1, \cdots, p)$ and let E^i be the aggregate of 2-simplexes of M'' that lie in the 2-cell of M^* dual to σ_1^i and are incident with $B^i(i=1, \cdots, p)$. The M''-neighborhood of T is a 3-cell and if we add to it the remainder of the M''-neighborhood of G excepting the 2-cells $E^1, \cdots, E^p$, we obtain a 3-cell. Hence R_1 is a canonical region.

By exactly the same argument it can be shown that R_2 is also a canonical region. It therefore follows that L is a canonical surface dividing the manifold M into the two canonical regions, R_1 and R_2, as we wished to prove.

7. We have proved incidentally that L and M are together orientable or non-orientable. Also, since the number of 2-cells which when removed reduce R_1 to a 3-cell is equal to the maximum number of non-intersecting circuits that can be drawn on L without disconnecting it, it follows that the number of 2-cells removed from R_1 is equal to the number removed from R_2. In other words, the cyclomatic number of G is equal to the cyclomatic number of G^*†.

8. We prove the following lemma:

LEMMA 10. *Let H and J be subcomplexes of the linear graph G of a manifold M such that it is possible to build up H from J by successively adding to J closed 1-simplexes of G in such fashion that at every step the 1-simplex which is being added either has one and only one end point in common with the subgraph already built up or else is the third side of a 2-simplex of which the other two sides*

† The cyclomatic number of a linear graph is the number of independent 1-circuits on it, i.e. the minimum number of 1-simplexes that can be removed which reduce the graph to a tree.

already belong to the subgraph. Then if the boundary of the M''-neighborhood of J is a canonical surface, so is the boundary of the M''-neighborhood of H, and conversely.

The proof of this lemma follows by induction. Suppose that at the nth step we have built up the subgraph J_n from J and that the boundary of the M''-neighborhood of J_n is a canonical surface of M. We add the closed 1-simplex σ_1^{n+1} to J_n, obtaining the subgraph J_{n+1}. By hypothesis, either σ_1^{n+1} has only one end point in common with J_n, or else it completes a triangle, of which the other two sides are already in J_n. It is obvious in the first case that the boundary of the M''-neighborhood of J_{n+1} is a canonical surface of M.

Let then σ_1^{n+1} complete a triangle of which the other two sides belong to J_n. Then if we call R_2 the M''-neighborhood of J_n, L its boundary and R_1 the remainder of M, that part of σ_1^{n+1} lying in R_1 can play the rôle of A of Lemma 2. Hence, applying Lemma 2, we obtain a new canonical surface of M which is precisely the boundary of the M''-neighborhood of J_{n+1}. It follows that the boundary of the M''-neighborhood of H is a canonical surface of M.

Conversely, let us suppose that the boundary of the M''-neighborhood of J_{n+1} is a canonical surface. Again, if the 1-simplex σ_1^{n+1} which was added to J_n to form J_{n+1} has only one end point in common with J_n, then the boundary of the M''-neighborhood of J_n is a canonical surface of M.

In the second case, let us call that part of the 2-simplex whose boundary σ_1^{n+1} completes lying in R_1, E, and F that part of the 2-cell dual to σ_1^{n+1} lying in R_2. We can choose E and F canonical 2-cells of R_1 and R_2, respectively, and so arrange that their boundaries (which meet once and only once) do not meet any of the other canonical curves. Lemma 1 is then applicable, and therefore the boundary of the M''-neighborhood of J_n is a canonical surface of M. It follows that the boundary of the M''-neighborhood of J is a canonical surface of M and the lemma is proved.

9. Let L and L' be the boundaries of the M''-neighborhoods of H and J, respectively, as above, and suppose that one (and hence the other) is a canonical surface. As an immediate consequence of Lemma 10 we have

COROLLARY 1. *Any special Heegaard diagram arising from L is equivalent to any special Heegaard diagram arising from L'.*

The proof follows at once from the fact that when we pass from J_n to J_{n+1}, we either do not change the Heegaard diagram at all, or else we modify it by a move of Type III.

10. At this point it becomes necessary to restrict the notion of homeomorphism. Let M and M_0 be two manifolds which are not only homeomorphic

but also equivalent in the sense of semi-linear analysis situs, i.e., it is possible to subdivide each of them so that the subdivisions have identical structures (cf. Alexander, J. W., *The combinatorial theory of complexes,* Annals of Mathematics, (2), vol. 31 (1930)). M and M_0 will then be called *equivalent* manifolds.

From this definition and from Lemma 10 and its corollary follows at once

THEOREM 5. *If M and M_0 are equivalent manifolds, G and G_0 their linear graphs, L and L_0 the boundaries of the M''- and M_0''-neighborhoods of G and G_0, respectively, then any special Heegaard diagram of M arising from L is equivalent to any special Heegaard diagram of M_0 arising from L_0.*

11. We have one more theorem to prove before we can prove our objective, Theorem 7.

THEOREM 6. *Let G be the linear graph of a manifold M, L_0 the boundary of the M''-neighborhood of G, and L any canonical surface whatsoever of M. Then any Heegaard diagram Δ of M arising from L is equivalent to any special Heegaard diagram Δ_0 arising from L_0.*

Let us note that if the canonical surface L is isotopic to a surface L', then L' is also a canonical surface of M.

Let L divide the manifold M into the canonical regions R_1 and R_2 and let $E_2^1, \cdots, E_2^p$ be a set of canonical 2-cells for R_1. After subdivision of M, if necessary, we can group the simplexes of M into cells, so that R_1 is composed of $p+1$ 3-cells, $E_3^1, \cdots, E_3^{p+1}$, where E_3^i is bounded by the 2-cell E_2^i and $E_2'^i$ (a 2-cell isotopic to E_2^i) and 2-simplexes on L, for $i=1, 2, \cdots, p$; and E_3^{p+1} is a 3-cell bounded by $\sum(E_2^i+E_2'^i)$ and simplexes on L. By elementary operations we can further change the structure of M so that all of the 3-cells $E_3^1, \cdots, E_3^{p+1}$ and all of the 2-cells $E_2^1, \cdots, E_2^p, E_2'^1, \cdots, E_2'^p$, become stars of simplexes having for centers the 0-cells $E_0^1, \cdots, E_0^{p+1}, P_0^1, \cdots, P_0^p$, and $P_0'^1, \cdots, P_0'^p$, respectively. The manifold M is thus transformed into an equivalent manifold N. Let L_n be the image of L in N.

Consider the linear graph H of N, where

$$H = \sum_{i=1}^{p}(E_0^{p+1}P_0^i + P_0^iE_0^i + E_0^iP_0'^i + P_0'^iE_0^{p+1})$$

in which $E_0^{p+1}P_0^i$, etc., stands for the 1-simplex joining E_0^{p+1} and P_0^i. By construction, G_n, the linear graph of N can be built up from H as is required in Lemma 10. Hence any special Heegaard diagram of N, say Δ_N, that arises from the boundary of the N''-neighborhood of G_n is equivalent to any special Heegaard diagram that arises from the boundary of the N''-neighborhood of

H. But this latter surface is isotopic to L_n. Hence Δ_N is equivalent to any Heegaard diagram arising from L_n. But since L_n is the image of L, Δ is such a diagram, hence Δ and Δ_N are equivalent. But by Theorem 5, Δ_N and Δ_0 are equivalent, therefore Δ and Δ_0 are equivalent as we wished to prove.

12. From Theorems 2, 3, 4, 5, and 6 follows the all inclusive theorem

THEOREM 7. *If Δ and Δ' are any two Heegaard diagrams whatsoever arising from a manifold M, then Δ and Δ' are equivalent.*

V. The general theorem

1. Several theorems follow immediately from those of Part IV.

THEOREM 8. *If the equivalent manifolds M and N give rise to the Heegaard diagrams Δ and Δ', then Δ and Δ' are equivalent.*

THEOREM 9. *If the diagram Δ gives rise to the manifold M, which in turn gives rise to the diagram Δ', then Δ and Δ' are equivalent.*

From Theorem 9 and the preceding theorems follows

THEOREM 10. *If the Heegaard diagrams Δ and Δ' give rise to equivalent manifolds, then Δ and Δ' are equivalent.*

THEOREM 11. *If the manifolds M and N give rise to equivalent diagrams, then M and N are equivalent.*

From Theorems 1, 8, 10, and 11 follows the general theorem

THEOREM 12. *Equivalent manifolds give rise to and arise from equivalent Heegaard diagrams and equivalent Heegaard diagrams give rise to and arise from equivalent manifolds.*

The problem of determining when two given 3-dimensional manifolds are equivalent (in the sense that we can pass from one to the other by elementary operations) is thus reduced to the problem of determining when two given Heegaard diagrams are equivalent (in the sense that we can pass from one to the other by means of the moves defined in Part III).

PRINCETON UNIVERSITY,
PRINCETON, N. J.

NON-CYCLIC ALGEBRAS OF DEGREE AND EXPONENT FOUR*

BY

A. ADRIAN ALBERT

1. Introduction. I have recently† proved the existence of non-cyclic normal division algebras. The algebras I constructed are algebras A of order sixteen (degree four, so that every quantity of A is contained in some quartic sub-field of A) containing no *cyclic* quartic sub-field and hence not of the cyclic (Dickson) type. But each A is expressible as a direct product of two (cyclic) algebras of degree two (order four). Hence the question of the existence of non-cyclic algebras *not* direct products of cyclic algebras, and therefore of essentially more complex structures than cyclic algebras, has remained unanswered.

The exponent of a normal division algebra A is the least integer e such that A^e is a total matric algebra. A normal division algebra of degree four has exponent two or four according as it is or is not expressible as a direct product of algebras of degree two.‡ I shall prove here that there exist non-cyclic normal division algebras of degree and exponent four, algebras of a more complex structure than any previously constructed normal division algebras.

2. Algebras of order sixteen. We shall consider normal simple algebras of order sixteen (degree four) over a field K. Algebra A has a quartic sub-field $K(u, v)$ where

$$u^2 = \rho, \quad v^2 = \sigma \qquad (\rho, \sigma \text{ in } K), \tag{1}$$

such that neither ρ, σ, nor $\sigma\rho$ is the square of any quantity of K. Algebra A contains quantities

$$j_1, j_2, j_3 = j_1 j_2,$$

such that

$$j_1 u = u j_1, \quad j_1 v = -v j_1, \quad j_1^2 = g_1 = \gamma_1 + \gamma_2 u \neq 0 \quad (\gamma_1, \gamma_2 \text{ in } K), \tag{2}$$

$$j_2 v = v j_2, \quad j_2 u = -u j_2, \quad j_2^2 = g_2 = \gamma_3 + \gamma_4 v \neq 0 \quad (\gamma_3, \gamma_4 \text{ in } K), \tag{3}$$

$$j_2 j_1 = \alpha j_3, \quad j_3^2 = g_3 = \gamma_5 + \gamma_6 uv \qquad (\gamma_5, \gamma_6 \text{ in } K), \tag{4}$$

* Presented to the Society, August 31, 1932; received by the editors June 9, 1932.

† In a paper published in the Bulletin of the American Mathematical Society, June, 1932. (Designated by Albert 1.)

‡ See Theorem 6 of my *Normal division algebras of degree four*, etc., these Transactions, vol. 34 (1932), pp. 363–372. (Designated by Albert 2.)

$$\alpha = \frac{\gamma_5 - \gamma_6 uv}{(\gamma_1 + \gamma_2 u)(\gamma_3 - \gamma_4 v)}. \tag{5}$$

A necessary and sufficient condition that A be associative is that

$$\gamma_5^2 - \gamma_6^2 \sigma\rho = (\gamma_1^2 - \gamma_2^2 \rho)(\gamma_3^2 - \gamma_4^2 \sigma). \tag{6}$$

A necessary and sufficient condition* that A be not expressible as a direct product of two algebras of degree two (that is, have exponent four) is that the equation

$$\alpha_1^2 - \alpha_2^2 \sigma - (\gamma_1^2 - \gamma_2^2 \rho)\alpha_3^2 = 0 \tag{7}$$

be impossible for any $\alpha_1, \alpha_2, \alpha_3$ not all zero and in K.

Algebra† A has a sub-algebra $B = (1, v, j_1, vj_1)$ over $K(u)$. This algebra is a generalized quaternion algebra and it is well known that B is a division algebra if and only if

$$g_1 \neq a_1^2 - a_2^2 \sigma \tag{8}$$

for any a_1 and a_2 in $K(u)$. But if $a_1 = \alpha_1 + \alpha_2 u$, $a_2 = \alpha_3 + \alpha_4 u$, the equation $g_1 = a_1^2 - a_2^2 \sigma$ implies that $\gamma_1 + \gamma_2 u = [\alpha_1^2 + \alpha_2^2 \rho - \sigma(\alpha_3^2 + \alpha_4^2 \rho)] + 2(\alpha_1\alpha_2 - \sigma\alpha_3\alpha_4)u$ so that $\gamma_1 = \alpha_1^2 + \alpha_2^2 \rho - \sigma(\alpha_3^2 + \alpha_4^2 \rho)$. We have now

THEOREM 1. *A sufficient condition that B be a division algebra is that the quadratic form*

$$Q = (\alpha_1^2 + \alpha_2^2 \rho) - \sigma(\alpha_3^2 + \alpha_4^2 \rho) - \gamma_1 \alpha_5^2 \tag{9}$$

in the variables $\alpha_1, \cdots, \alpha_5$ *shall not vanish for any* $\alpha_1, \cdots, \alpha_5$ *not all zero and in* K.

For if the sufficient condition of Theorem 1 were satisfied and yet B were not a division algebra we would have $\gamma_1 = \alpha_1^2 + \alpha_2^2 \rho - \sigma(\alpha_3^2 + \alpha_4^2 \rho)$ so that $Q = 0$ for $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ in K and $\alpha_5 = 1$, a contradiction.

It is also known‡ that, when B is a division algebra, A is also a division algebra if and only if there is no quantity X in B for which

$$g_2 = X'X, \tag{10}$$

where if $X = b + dj_1$ then $X' = b(-u) + d(-u)\alpha j_1$ with a and b of course in $K(u, v)$.

* See Albert 2.

† For the properties of this section see my paper in these Transactions, vol. 32 (1930), pp. 171–195. (Designated hereafter by Albert 3.)

‡ See L. E. Dickson's *Algebren und ihre Zahlentheorie*, p. 64, for both the condition that B be a division algebra and A be a division algebra.

I have proved* that

$$(bj_2)^2 = f_3 + f_4 v, \qquad (dj_3)^2 = f_5 + f_6 uv, \tag{11}$$

where if

$$b = \beta_1 + \beta_2 v + (\beta_3 + \beta_4 v)u, \qquad d = \delta_1 + \delta_2 uv + (\delta_3 + \delta_4 uv)u \tag{12}$$

and

$$b_1 = \beta_1^2 + \beta_2^2\sigma - \rho(\beta_3^2 + \beta_4^2\sigma), \qquad b_2 = 2(\beta_1\beta_2 - \rho\beta_3\beta_4), \tag{13}$$

$$d_1 = \delta_1^2 + \delta_2^2\sigma\rho - \rho(\delta_3^2 + \delta_4^2\sigma\rho), \qquad d_2 = 2(\delta_1\delta_2 - \sigma\rho\delta_3\delta_4), \tag{14}$$

then

$$\begin{aligned} f_3 &= b_1\gamma_3 + b_2\sigma\gamma_4, & f_4 &= b_1\gamma_4 + b_2\gamma_3, \\ f_5 &= d_1\gamma_5 + d_2\sigma\rho\gamma_6, & f_6 &= d_1\gamma_6 + d_2\gamma_5. \end{aligned} \tag{15}$$

I have also shown that if $g_2 = X'X$ then

$$f_4 = f_6 = 0, \quad f_3 + f_5 = \gamma_3^2 - \gamma_4^2\sigma. \tag{16}$$

But then $\gamma_3 b_2 = -\gamma_4 b_1$, $\gamma_5 d_2 = -\gamma_6 d_1$, so that from (16_2), (15),

$$\gamma_3\gamma_5(\gamma_3^2 - \gamma_4^2\sigma) = (\gamma_3^2 - \gamma_4^2\sigma)\gamma_5 b_1 + (\gamma_5^2 - \gamma_6^2\sigma\rho)\gamma_3 d_1. \tag{17}$$

If A is associative then (6) is satisfied. Also $g_2 \neq 0$ so that $g_2(-v) \neq 0$, $\gamma_3^2 - \gamma_4^2\sigma \neq 0$. Then (17) is equivalent to

$$\gamma_3\gamma_5 = \gamma_5 b_1 + \gamma_3 d_1(\gamma_1^2 - \gamma_2^2\rho). \tag{18}$$

As in the proof of Theorem 1 we have immediately

THEOREM 2. *A sufficient condition that A with division sub-algebra B be a division algebra is that the quadratic form*

$$\begin{aligned} Q \equiv \gamma_5[(\alpha_1^2 + \alpha_2^2\sigma) - \rho(\alpha_3^2 + \alpha_4^2\sigma)] \\ + \gamma_3(\gamma_1^2 - \gamma_2^2\rho)[(\alpha_5^2 + \alpha_6^2\sigma\rho) - \rho(\alpha_7^2 + \alpha_8^2\sigma\rho)] - \gamma_3\gamma_5\alpha_9^2 \end{aligned} \tag{19}$$

shall not vanish for any $\alpha_1, \cdots, \alpha_9$ *not all zero and in K.*

3. **Algebras over $K(q)$.** Let $L = K(q)$ be a quadratic field over K where

$$q^2 = \delta = \delta_1^2 + \delta_2^2 \qquad (\delta_1 \text{ and } \delta_2 \text{ in } K). \tag{20}$$

It is well known that if K contains no quantity k such that $k^2 = -1$ then every cyclic quartic field over K contains a quadratic sub-field L of the above type. Hence a sufficient condition that an algebra of degree four be non-cyclic is that A contain no quadratic sub-field L as above. But also A contains no sub-

* Albert 3, p. 178.

field equivalent to any given quadratic field L if and only if $A \times L$ is a division algebra.* Hence we have

THEOREM 3. *If no k in K has the property $k^2 = -1$, a sufficient condition that a normal simple algebra A of order sixteen over K be a non-cyclic normal division algebra is that $A \times L$ be a division algebra for every quadratic field $L = K(q)$,*

$$q^2 = \delta = \delta_1^2 + \delta_2^2 \qquad (\delta_1 \text{ and } \delta_2 \text{ in } K). \tag{21}$$

We shall apply Theorem 3 as follows. We shall choose a particular field of reference, K. We shall then define A by a choice of $\rho, \sigma, \gamma_1, \cdots, \gamma_6$. Then also $A \times L$ is evidently a normal simple algebra (of the same kind as A over K) over L when we show that neither ρ, σ, nor $\sigma\rho$ is the square of any quantity of L (not merely K). We shall then prove that A(not $A \times L$ which can have exponent two) has exponent four, while $A \times L$ is a division algebra. This latter step will be an application of Theorems 1 and 2 applied to $A \times L$ over L. The algebras A over K will be non-cyclic algebras of exponent four by Theorem 3.

4. **The field K.** Let F be any *real number* field, and let x, y, and z be independent marks (indeterminates). The field $F(x, y, z) \equiv K$ is a function field consisting of all rational functions with (real) coefficients in F of x, y, z. We shall deal with quadratic forms Q and equations $Q = 0$ so that we shall always be able to delete denominators and hence take our quantities in

$$J = F[x, y, z],$$

the domain of integrity consisting of all polynomials in x, y, z with coefficients in F. We shall of course also consider the domains $F[x]$, $F[x, y]$, etc.

Consider a field $K(q)$ as in §3. It is evident that the quantity q defining such a quadratic field may always be chosen so that $\delta, \delta_1, \delta_2$ are in J. Also in a quadratic form $Q = 0$ with coefficients in J and variables over $K(q)$ we may always take the variables to be in the domain of integrity $J[q]$ of all quantities of the form

$$a + bq$$

where a and b are in J.

Every quantity $a = a(x, y, z)$ of J has a highest power z^n with coefficient in $F[x, y]$ not identically zero. We shall call n the z-degree of a, the coefficient of z^n the *z-leading coefficient* of a. Similarly a has an x-degree, y-degree, x-leading coefficient, y-leading coefficient. A restriction of the z-degree of a certain expression and its z-leading coefficient evidently does not affect its x-degree, etc.

* Cf. Albert 1.

If the coefficient of z^n above is $b(y, x)$ and the coefficient of the highest power y^m of y in b is $c(x)$, then m is called the (z, y)-degree of a, $c(x)$ the (z, y)-leading coefficient of a. Finally the degree of $c(x)$ is the (z, y, x)-degree of a, its leading coefficient in F, the (z, y, x)-leading coefficient of a.

We have similarly (x, y, z)-degree and leading coefficient, etc. Using these definitions an elementary result is

LEMMA 1. *The field K contains no quantity k such that $k^2 = -1$.*

For let $k^2 = -1$. Then $rk = s$, where r and s are in J and are both not zero. It follows that $s^2 = -r^2$. The (x, y, z)-leading coefficient of s^2 is evidently a real square and is positive, that of $-s^2$, negative so that the polynomial identity $r^2 = -s^2$ is impossible.

LEMMA 2. *There exist quantities λ, μ in $F[x, y]$ such that $\lambda^2+\mu^2$ is not the square of any quantity of $F(x, y)$.*

We prove the above lemma with the example $\lambda = x$, $\mu = y$. If $x^2+y^2 = b^2$, where b is a rational function of x and y, it is evident that b must be a polynomial in x and y. For the square of a rational function in its lowest terms and with denominator not unity is never a polynomial. Hence we may put $b = b_1x + b_2$ where b_2 is in $F[y]$, b_1 merely in $F[x, y]$. Then $x^2+y^2 = b_1^2x^2 + 2b_1b_2x + b_2^2$ identically in x and y. It follows that $b_2^2 = y^2$, $b_2 = \pm y$. Then $x^2 = b_1^2x^2 \pm 2b_1xy$. Hence b_1 divides x and is a power of x. But then $\pm(2b_1)y = x - b_1^2x$ in $F[x]$, b_1 in $F(x)$, which is impossible.

5. **The S-polynomials.** The quadratic forms (9), (19) over L shall be treated as follows. If $Q = \sum \alpha_i^2\lambda_i$ with λ_i in J (not in $J[q]$) vanishes for α_i in L and not all zero, then obviously, by multiplying Q by the square of the least common denominator, not zero and in J, of the $\alpha_i = \alpha_{i1} + \alpha_{i2}q$ (α_{i1}, α_{i2} in K), we shall have $Q = 0$ for α_i in $J[q]$, that is, α_{i1} and α_{i2} in J. But then

$$Q = \sum \lambda_i[(\alpha_{i1}^2 + \alpha_{i2}^2\delta) + (2\alpha_{i1}\alpha_{2i})q] = 0$$

so that

$$\sum \lambda_i S_i = 0,$$

where

$$S_i = (\alpha_{i1})^2 + (\alpha_{i2}\delta_1)^2 + (\alpha_{i2}\delta_2)^2. \tag{22}$$

We shall call a polynomial of the form (22) an S-polynomial. All such polynomials have the properties that all their degrees are even, all their (, ,)-leading coefficients positive. Moreover such a polynomial is zero if and only if $\alpha_i = \alpha_{i1} = \alpha_{i2} = 0$. Hence we have

LEMMA 3. *A sufficient condition that a quadratic form* $\sum \lambda_i \alpha_i^2$ *with* λ_i *in* J *shall not vanish for any* α_i *not all zero and in* $K(q)$ *is that* $\sum \lambda_i S_i$ *shall not vanish for any S-polynomials* S_i *not all zero.*

6. **The multiplication constants of** A. We now choose $\rho, \sigma, \gamma_1, \cdots, \gamma_6$ in J. We shall take

(23) σ *of even z-degree, even* (z, y)*-degree, odd* (z, y, x)*-degree.*

We shall define γ_1 and γ_5 in terms of certain quantities ϵ_1, ϵ_5, where

(24) (*the z-degree of* ϵ_5 *is odd*) $>$ (*z-degree of* $\epsilon_1\gamma_3$);

(25) (*the z-degree of* γ_3 *is odd*) $>$ (*z-degree of* $\gamma_4\sigma$);

(26) (*the z-degree of* γ_2) $>$ (*z-degree of* $\gamma_6\sigma$);

(27) *the* (z, y)*-degree of* γ_3 *even, of* ϵ_5 *odd.*

The above conditions are restrictions merely on the z-leading coefficients of our quantities. By making the corresponding z-degrees sufficiently large we evidently only restrict a single term in each quantity, satisfy the above conditions, and yet permit any desired inequalities between x-degrees, y-degrees of the same quantities. Moreover (, ,)-leading coefficients other than the (z, ,)-leading coefficients may be taken to have any desired sign, and the evenness or oddness of (,)-degrees, etc., other than those already given above are still at our choice. We therefore may continue with

(28) σ *of even y-degree, odd* (y, x)*-degree;*

(29) (*y-degree of* ϵ_1 *odd*) $>$ (*y-degree of* ϵ_5);

(30) (*y-degree of* γ_2) $>$ (*y-degree of* $\gamma_6\sigma$);

(31) (*y-degree of* γ_3) $>$ (*y-degree of* $\gamma_4\sigma$);

(32) σ *of odd x-degree.*

Let the x-leading coefficient of γ_6 be π_1, that of $\gamma_2\gamma_4$ be π_2 such that

(33) $\pi_1^2 + \pi_2^2 \neq \lambda^2$ *for any* λ *of* $F(y, z)$.

This restriction may be satisfied by Lemma 2 and there merely restricts the x-leading coefficients of γ_6 and $\gamma_2\gamma_4$. Also take

(34) (*x-degree of* γ_6) $=$ (*x-degree of* $\gamma_2\gamma_4$) $>$ (*x-degree of* $\gamma_2\gamma_3$),

that is, the x-degree of γ_4 greater than the x-degree of γ_3, and, if we desire, the x-leading coefficient of γ_2 unity, that of γ_4, y, that of γ_6, z, and (33) is satisfied.

Finally let

$$e = \gamma_2^2(\gamma_3^2 - \gamma_4^2\sigma) - \gamma_6^2\sigma, \tag{35}$$

$$\rho = e[\epsilon_1^2(\gamma_3^2 - \gamma_4^2\sigma) - \epsilon_5^2], \tag{36}$$

$$\gamma_1 = \epsilon_1 e, \quad \gamma_5 = \epsilon_5 e. \tag{37}$$

Then

$$\begin{aligned}\gamma_1^2 - \gamma_2^2\rho &= \epsilon_1^2 e^2 - \gamma_2^2\rho \\ &= e\epsilon_1^2[\gamma_2^2(\gamma_3^2 - \gamma_4^2\sigma) - \gamma_6^2\sigma] - e\gamma_2^2\epsilon_1^2(\gamma_3^2 - \gamma_4^2\sigma) + \gamma_2^2\epsilon_5^2 e,\end{aligned}$$

and

$$\gamma_1^2 - \gamma_2^2\rho = e[(\gamma_2\epsilon_5)^2 - (\gamma_6\epsilon_1)^2\sigma]. \tag{38}$$

Also

$$\begin{aligned}\gamma_5^2 - \gamma_6^2\sigma\rho &= \epsilon_5^2 e^2 - \gamma_6^2\sigma\rho \\ &= e\gamma_2^2\epsilon_5^2(\gamma_3^2 - \gamma_4^2\sigma) - e\gamma_6^2\epsilon_5^2\sigma + e\gamma_6^2\sigma\epsilon_5^2 - e\gamma_6^2\sigma\epsilon_1^2(\gamma_3^2 - \gamma_4^2\sigma) \\ &= (\gamma_3^2 - \gamma_4^2\sigma)e[(\gamma_2\epsilon_5)^2 - (\gamma_6\epsilon_1)^2\sigma].\end{aligned}$$

By (38) we have

THEOREM 4. *If $\rho, \sigma, \gamma_1, \cdots, \gamma_6$ are chosen as in* (35), (36), (37), *the corresponding algebra A satisfies*

$$\gamma_5^2 - \gamma_6^2\sigma\rho = (\gamma_1^2 - \gamma_2^2\rho)(\gamma_3^2 - \gamma_4^2\sigma) \tag{39}$$

and is associative.

7. **Elementary properties.** In (25) we chose the z-degree of γ_3 to be greater than the z-degree of $\gamma_4\sigma$. In (26) we took the z-degree of γ_2 greater than that of $\gamma_6\sigma$. It now follows that the only term of e containing its highest power of z is $(\gamma_2\gamma_3)^2$. Similarly, by (24), (25) the term of $[\epsilon_1^2(\gamma_3^2 - \gamma_4^2\sigma) - \epsilon_5^2]$ containing its highest power of z is $-\epsilon_5^2$. Hence the term of ρ containing its highest power of z is $-(\gamma_2\gamma_3\epsilon_5)^2$.

LEMMA 4. *The z-degree of ρ is positive, even, and the z-leading coefficient of ρ is the negative of a perfect square.*

Consider the y-degree of ρ. By (31) the y-degree of $\gamma_3^2 - \gamma_4^2\sigma$ is positive and its y-leading coefficient is a perfect square (in γ_3^2). By (35) the leading y-term of e is then in $(\gamma_2\gamma_3)^2$, while the leading y-term of $\epsilon_1^2(\gamma_3^2 - \gamma_4^2\sigma) - \epsilon_5^2$ is then in $(\epsilon_1\gamma_3)^2$. Hence the term of ρ containing its highest power of y is $(\epsilon_1\gamma_2\gamma_3^2)^2$.

LEMMA 5. *The y-degree of ρ is positive and even, and its y-leading coefficient is a perfect square.*

Consider the x-degree of e. We have taken the x-degree of γ_6 equal to the x-degree of $\gamma_2\gamma_4$ and the x-degree of γ_4 greater than the x-degree of γ_3. But $e=-[(\gamma_2\gamma_4)^2+\gamma_6^2]\sigma+(\gamma_2\gamma_3)^2$. Hence the x-leading coefficient of e is the product of the x-leading coefficient of $-\sigma$ by $\pi_1^2+\pi_2^2$. But the x-degree of σ has been taken odd.

LEMMA 6. *Let σ_0 be the x-leading coefficient of σ. Then the x-leading coefficient of e is $-\sigma_0(\pi_1^2+\pi_2^2)$ and the x-degree of e is a positive odd integer.*

The quantity $\gamma_1^2-\gamma_2^2\rho$ is determined by (38). We shall require

LEMMA 7. *The z-degrees of $\gamma_1^2-\gamma_2^2\rho$ are all even.*

For proof we notice that we have already shown that the z-degree of e is even, in fact the leading term of e when arranged according to powers of z is a perfect square. Also we have taken the z-degree of $(\gamma_2\epsilon_5)^2$ greater than that of $(\gamma_6\epsilon_1)^2\sigma$. Hence the z-degree of $\gamma_1^2-\gamma_2^2\rho$ is even. In fact its z-leading coefficient occurs only in $(\gamma_2^2\epsilon_5\gamma_3)^2$ and is a perfect square, so that all its z-degrees are even.

One of the properties required in our definition of A is that neither ρ, σ, nor $\sigma\rho$ shall be the square of any quantities of K. We shall prove

LEMMA 8. *Neither ρ, σ, nor $\sigma\rho$ is the square of any quantity of $K(q)$.*

For let $\rho=\alpha^2$ where α is in $K(q)$. Then $\mu\alpha=\lambda$ where λ is in $J[q]$ and μ is in J. Then $\rho\mu^2=\lambda^2$ in J. A quantity λ of $K(q)$ has its square in K if and only if it is either in K or a multiple of q by a quantity of k. If λ in $J[q]$ is in K then λ is in J so that $\rho\mu^2=\lambda^2$ is impossible because the (z, y, x)-leading coefficient of ρ and hence $\rho\mu^2$ is negative while that of λ^2 is positive. Hence $\lambda=\nu q$ with ν in J. Then $\lambda^2=\nu^2\delta$ is an S-polynomial and cannot be identical with $\rho\mu^2$ of negative (z, y, x)-leading coefficient.

Similarly $\sigma\neq\alpha^2$ where we now use the property that σ has odd x-degree. Finally by (28) and Lemma 5 $\sigma\rho$ has odd (y, x)-degree and $\sigma\rho\neq\alpha^2$ for any α of $K(q)$.

COROLLARY 1. *The quantities ρ, σ, $\sigma\rho$ are not the squares of any quantities of K.*

It follows from Corollary 1 that $K(u, v)$ is a quartic field over K and that $g_1=0$ if and only if $\gamma_1=\gamma_2=0$. By Lemma 7, $g_1\neq 0$. Also (31) implies that $g_2\neq 0$, while the associativity condition (38) implies that $g_3\neq 0$.

8. **The exponent of A.** We shall use (7) to prove that A has exponent four, that is, A is not a direct product of two algebras of degree two. Assume that A has not exponent four so that (7) is satisfied for $\alpha_1, \alpha_2, \alpha_3$ in K and not all zero. As we have already remarked we may take $\alpha_1, \alpha_2, \alpha_3$ in J. If $\alpha_2=\alpha_3=0$,

$$(7) \qquad \alpha_1^2 - \alpha_2^2\sigma = (\gamma_1^2 - \gamma_2^2\rho)\alpha_3^2$$

implies that $\alpha_1^2 = \alpha_1 = 0$, a contradiction. Hence if $\alpha_3 = 0$ then $\alpha_2 \neq 0$ and $\sigma = (\alpha_1\alpha_2^{-1})^2$, a contradiction of Corollary 1. Thus $\alpha_3 \neq 0$.

By Lemma 7 $\gamma_1^2 - \gamma_2^2\rho \neq 0$ so that $h = (\gamma_2\epsilon_5)^2 - (\gamma_6\epsilon_1)^2\sigma \neq 0$. The equation $\gamma_1^2 - \gamma_2^2\rho = he$ gives

$$(\alpha_1^2 - \alpha_2^2\sigma)h = (\alpha_3 h)^2 e.$$

Let $\beta_3 = \alpha_3 h \neq 0$, $\beta_1 = \alpha_1\gamma_2\epsilon_5 + \alpha_2\gamma_6\epsilon_1\sigma$, $\beta_2 = \alpha_1\gamma_6\epsilon_1 + \alpha_2\gamma_2\epsilon_5$. Then, as may be easily computed,*

$$(40) \qquad \beta_1^2 - \beta_2^2\sigma = e\beta_3^2 \qquad (\beta_3 \neq 0, \beta_1, \beta_2, \beta_3 \text{ in } J).$$

But then $\beta_1^2 = \sigma\beta_2^2 + e\beta_3^2$. The x-leading coefficient of $e\beta_3^2$ has the form $-\sigma_0(\pi_1^2 + \pi_2^2)\beta_{30}^2$ by Lemma 6. The x-leading coefficient of $\sigma\beta_2^2$ has the form $\sigma_0\beta_{20}^2$. But $(\pi_1^2 + \pi_2^2)\beta_{30}^2 \neq 0$ is not the square of any quantity of $K(y, z)$. Hence the x-leading coefficient of $\sigma\beta_2^2 + e\beta_3^2$ is not zero. But the x-degree of this expression is odd since σ has odd x-degree, e has odd x-degree, $\beta_3 \neq 0$. It follows that (40) is impossible for $\beta_3 \neq 0$, a contradiction.

9. **The first norm condition.** We wish to prove that algebra B is a division algebra, that is, prove that $g_1 \neq a \cdot a(-v)$ for any a of $K(u, v)$, the so called *first norm condition*. As we have shown this condition will be satisfied if we can show that the equation

$$(41) \qquad S_1 + S_2\rho - \sigma(S_3 + S_4\rho) = \gamma_1 S_5$$

is impossible for S-polynomials $S_1, \cdots, S_5$ not all zero, a consequence of §5 applied to (9).

By Lemma 2 the y-degree of ρ is even and the (y, z, x)-leading coefficient of ρ is positive. Also the y-degree of σ is even. Hence the y-degree of each of $S_1, S_2\rho, S_3, S_4\rho$ is even. But the (y, z, x)-leading coefficients of these terms are all positive. Moreover $S_1 + S_2\rho$, $S_3 + S_4\rho$ have even (y, z)-degree, while σ has odd (y, z)-degree. Hence the (y, z)-degree of $S_1 + S_2\rho - \sigma(S_3 + S_4\rho)$ is either even or odd according as the (y, z)-degree of $S_1 + S_2\rho$ is greater or less than the (y, z)-degree of $(S_3 + S_4\rho)\sigma$. In any case the corresponding (y, z, x)-leading coefficient is zero if and only if $S_1 = S_2 = S_3 = S_4 = 0$. We have shown that $T = S_1 + S_2\rho - \sigma(S_3 + S_4\rho)$ has even y-degree and (y, z, x)-leading coefficient zero if and only if $S_i = 0$ $(i = 1, \cdots, 4)$.

By (35), (30), (31) the y-degree of e is even. By (37), (29) the y-degree of γ_1 is odd. Hence the y-degree of $\gamma_1 S_5$ is odd unless $S_5 = 0$. But $\gamma_1 S_5 = T$ has even y-degree. Hence $S_5 = 0$, $T = 0$, T has (y, z, x)-leading coefficient zero so that $S_i = 0$ $(i = 1, \cdots, 5)$.

* That is, let $a = \alpha_1 + \alpha_2 v$, $b = \gamma_2\epsilon_5 + \gamma_6\epsilon_1 v$. Then $ab = (\alpha_1\gamma_2\epsilon_5 + \alpha_2\gamma_6\epsilon_1\sigma) + (\alpha_1\gamma_6\epsilon_1 + \alpha_2\gamma_2\epsilon_5)v = \beta_1 + \beta_2 v$, and $a \cdot a(-v) \cdot b \cdot b(-v) = (\alpha_1^2 - \alpha_2^2\sigma) \cdot h = ab \cdot \overline{ab}(-v) = \beta_1^2 - \beta_2^2\sigma$.

10. The second norm condition. This is the condition $g_2 = X'X$ which, by §5 and (19), is satisfied if we can prove that

$$\gamma_5[S_1+S_2\sigma-\rho(S_3+S_4\sigma)]+\gamma_3(\gamma_1^2-\gamma_2^2\rho)[S_5+S_6\sigma\rho-\rho S_7-\sigma S_8]=\gamma_3\gamma_5 S_9 \tag{42}$$

is impossible for S-polynomials $S_i (i=1, \cdots, 9)$ not all zero. Notice that we have replaced $\rho\alpha_8^2\rho=(\rho\alpha_8)^2$ of (19) by the S-polynomial S_8 instead of the formally corresponding $\rho^2 S_8$.

By (24) the z-degree of γ_3 is odd. By the proof of Lemma 4 the z-degree of e is even and the z-leading coefficient of e is a perfect square. Applying (27) we have

LEMMA 9. *The z- and (z, y)-degrees of γ_5 are odd.*

We have taken ρ to have all even degrees and *negative* (z, y, x)-leading coefficient by Lemma 4. Also σ has even z-degree, (z, y)-degree, but odd (z, y, x)-degree. Hence the (z, y, x)-leading coefficient of any $S_i-\rho S_j$ is positive or zero according as not both or both of S_i, S_j are zero. Hence the (z, y, x)-leading coefficient of a combination $T=S_i-\rho S_j\pm\sigma(S_r-\rho S_t)$ is zero if and only if the four S_i are zero. Moreover T has even (z, y)-degree and (z, y)-leading coefficient which is identically zero only when all the four S_i are zero. But the (z, y)-degree of γ_3 is even, the (z, y)-degree of $\gamma_1^2-\gamma_2^2\rho$ is even, while that of γ_5 is odd. Hence the (z, y)-leading coefficient of

$$R=\gamma_5[(S_1-\rho S_3)+\sigma(S_2-\rho S_4)]+\gamma_3(\gamma_1^2-\gamma_2^2\rho)[S_5-\rho S_7-\sigma(S_6-\rho S_8)]$$

is either the (z, y)-leading coefficient of its first bracket or of its second bracket, while R has z-leading coefficient identically zero if and only if $S_i=0$ $(i=1, \cdots, 8)$. But the z-degree of R is *odd* unless the S_i are zero since the z-degree of γ_3 is odd by (25), that of γ_5 odd by Lemma 9. By (42) $R=\gamma_3\gamma_5 S_9$ has *even* z-degree. Hence $R=0$, $S_9=0$, and R has z-leading coefficient zero. This proves that $S_i=0$ $(i=1, \cdots, 9)$ as desired. We have proved

LEMMA 10. *Let F be a real number field, x, y, z indeterminates, and let A be an algebra of order sixteen over $K=F(x, y, z)$ defined by* (1)–(5), (23)–(37). *Then A is a normal division algebra of degree and exponent four over K, $A\times L$ is a normal division algebra of degree four over L for every quadratic field $L=K(q)$, $q^2=\delta=\delta_1^2+\delta_2^2$ (δ_1, δ_2 in K).*

As an immediate corollary of Lemma 10 we then have

THEOREM. *The algebras of Lemma* 10 *are non-cyclic algebras of degree four not expressible as direct products of cyclic algebras of degree two.*

UNIVERSITY OF CHICAGO,
CHICAGO, ILL.

ON ABELIAN FIELDS*

BY

LEONARD CARLITZ†

1. Introduction

By Kronecker's‡ Theorem on Abelian fields, all such fields are subfields of cyclotomic fields, that is, fields generated by a root of unity. Abelian fields may then be classified by considering all cyclotomic fields and sorting the subfields in some manner that will exclude repetition. For example this is done, in part at least, by Weber by making use of the notion of *primary* subfields: a subfield of Ω_m, the field generated by a primitive mth root of unity, is a primary subfield if it is not contained in an $\Omega_{m'}$ $(m'<m)$. We here make use of what we shall call *simple*§ (primary) subfields as defined below. If then the (known) discriminants of Abelian fields are set up on this basis, a number of properties of Abelian fields become apparent. In particular is this true of the fields contained in a fixed simple subfield (see §5).

In §6 some results on common index divisors (that is, common inessential discriminantal divisors) are obtained. Using a necessary and sufficient condition valid for any algebraic field it is shown how to derive for the case of Abelian fields very simple criteria that a given rational prime be a common index divisor. The criteria are of two kinds. A typical instance of the first kind is the following.

Let q and l be odd primes such that $l\equiv 1 \pmod q$; let C denote that cyclic subfield of Ω_l that is of degree q. Then a necessary and sufficient condition that a prime $p(p<q)$ be a common index divisor of C is that

$$p^{(l-1)/q} \equiv 1 \pmod{q}.$$

As an instance of the criteria of the second kind, we quote the following theorem:

Let K be Abelian of degree q^n and type $(1, 1, \cdots)$. *Then if d is the discriminant of K, and if*

---

* Presented to the Society, November 29, 1930; received by the editors March 11, 1932.

† International Research Fellow.

‡ See Hilbert, *Die Theorie der algebraischen Zahlkörper*, Jahresbericht der Deutschen Mathematiker Vereinigung, vol. 4 (1894–1895), Theorem 131.

§ Called "Ausgangs-Kreiskörper" by M. Gut, *Die Zetafunktion, die Klassenzahl und die Kronecker'sche Grenzformel eines beliebigen Kreiskörpers*, Commentarii Mathematici Helvetici, vol. 1 (1929), p. 160.

(i) p *does not divide* d, $p \leqq q^{n/q}$;

(ii) $p \mid d$,* $p \leqq q^{(n-1)/q}$,

p is surely a common index divisor of K.

We shall suppose in what follows that all the discriminants are *odd* unless the contrary is explicitly stated; this makes for a considerable simplification and avoids listing a great many exceptional cases.

2. Classification

Let m be an integer $\geqq 3$, and let Ω_m be the field defined by a primitive mth root of unity. We suppose the group† G_ϕ of Ω_m exhibited by a reduced residue system (mod m); ϕ stands for $\phi(m)$. Let m be divisible by exactly n (odd) primes:

$$m = q_1^{f_1} \cdots q_n^{f_n};$$

put

$$\phi_i = \phi(q_i^{f_i}) = q_i^{f_i-1}(q_i - 1) \qquad (i = 1, \cdots, n).$$

Let r_i' denote fixed primitive roots (mod $q_i^{f_i}$), respectively, and let r_i be defined (mod m) by

$$(1) \qquad \begin{aligned} r_i &\equiv r_i' \quad (\text{mod } q_i^{f_i}), \\ r_i &\equiv 1 \quad (\text{mod } q_j^{f_j}) \qquad (i \neq j). \end{aligned}$$

Then G_ϕ is generated by $r_1, \cdots, r_n$:

$$(2) \qquad G_\phi = \{r_1, \cdots, r_n\}.$$

We now define a *simple* (primary) subfield of Ω_m as one corresponding‡ to a group

$$(3) \qquad G_\mu = \{r_1^{\nu_1}, \cdots, r_n^{\nu_n}\},$$

where

$$(4) \qquad \phi_i = \mu_i\nu_i, \quad q_i \text{ does not divide } \mu_i \qquad (i = 1, \cdots, n),$$

G_μ is evidently of order $\mu = \mu_1 \cdots \mu_n$, and K, the field corresponding to G_μ, is of degree $\nu = \nu_1 \cdots \nu_n$. That K is indeed primary follows from the second part of equations (4).

It is now a simple matter to exhibit our mode of classification. We notice to begin with that any primary subfield k of Ω_m is contained, properly or

* As usual, read for $a \mid b$, "a divides b."
† Hilbert, loc. cit , p. 248.
‡ Hilbert, loc. cit., p. 250.

improperly, in a *unique* minimal simple subfield of Ω_m; for it is clear from (2) that the greatest common subfield of two simple subfields is itself a simple subfield of Ω_m. Let us then fix our attention on a particular simple subfield K. Choose any maximal subfield k_1 of K. If k_1 is primary (with respect to Ω_m), choose k_2, some maximal subfield of k_1; we continue this process until we arrive either at another simple subfield or else at a k_j none of whose subfields is primary. To illustrate the process, let us classify the primary subfields of Ω_m, $m=5^3\cdot 11^2$.

Let r_1, r_2 appertain to $5^2\cdot 4$, $11\cdot 10$, respectively (see (1) above). Then, if the group generated by $A, B, \cdots$ be denoted by $\{A, B, \cdots\}$, we get among others the following chains:

I. $\{1\}\sim\Omega_m$,

$$\{r_1^{50} r_2^{55}\} \sim k_1 \text{ of degree } \phi(m)/2,$$

$$\{r_1^{50}, r_2^{55}\} \sim k_2 \text{ (simple) of degree } \phi(m)/4.$$

II. Starting with k_2 we may choose one of

$$\{r_1^{50}, r_2^{55}, r_1^{10} r_2^{11i}\} \sim k_{2i} \text{ of degree } \phi(m)/20 \qquad (i = 1, \cdots, 4),$$

or

$$\{r_1^{25}, r_2^{55}\} \sim k_3 \text{ (simple) of degree } \phi(m)/8; \text{ etc.}$$

III. Starting with k_3, we may choose one of

$$\{r_1^{25}, r_2^{55}, r_1^{5} r_2^{11i}\} \sim k_{3i} \text{ of degree } \phi(m)/40 \qquad (i = 1, \cdots, 4);$$

no subfield of any k_{3i} is primary.

IV. In place of k_1 (of I) we may take

$$\{r_1^{20} r_2^{22i}\} \sim k_{1i}' \text{ of degree } \phi(m)/5 \qquad (i = 1, \cdots, 4),$$

$$\{r_1^{20}, r_2^{22}\} \sim k_2' \text{ (simple) of degree } \phi(m)/25,$$

and contained in each k_{1i}'.

These four chains will suffice to indicate how the classification may be carried out in any special case; the utility of this method of arrangement will appear below.

3. The discriminants

The form of the discriminant of an Abelian field is known, at least in the sense that the discriminant of any subfield of an Ω_m can be explicitly written

down.* As the explicit expression for the discriminants will be required they will be stated here in the form of lemmas. It is convenient, and indeed leads to an important result, first to calculate the discriminant of an arbitrary simple field, and then proceed to the case of an entirely arbitrary subfield.

LEMMA 1. *The discriminant of a simple primary subfield K of Ω_m is determined by*

$$d(K) = \pm \prod_{i=1}^{n} q_i^{t_i} \tag{5}$$

where

$$t_i = \frac{1}{\mu}\left(\frac{\phi s_i}{\phi_i} - (\mu_i - 1)\frac{\mu\nu}{\mu_i\nu_i}\right),$$

$$s_i = q_i^{f_i-1}(f_i(q_i - 1) - 1), \tag{6}$$

and $\mu_i, \nu_i, \phi_i, \mu, \nu$ are defined by (4).

If now k is any primary subfield of Ω_m, then as seen in §2 it either is itself simple or else is contained in a unique minimal simple subfield. Calling this field K, and assuming all the above notation for a simple field, we get

LEMMA 2. *The relative discriminant of K with respect to k is the unit ideal of k.*

Now by a general theorem†

$$d(K) = d^{\rho}(k)N(D),$$

where ρ is the relative degree of K/k, D is the relative discriminant of K/k, and $N(D)$ denotes the norm in k. Hence Lemmas 1 and 2 immediately imply

LEMMA 3. *The discriminant of an arbitrary primary k is determined by*

$$d(k) = d^{1/\rho}(K) = \pm \prod_{i=1}^{n} q_i^{t_i/\rho},$$

where t_i is defined by (6).

4. THE SUBFIELDS OF A SIMPLE SUBFIELD

Let us fix some K, a simple subfield of Ω_m, defined by equations (3) and (4), say. We shall consider the set of fields $\{k\}$ satisfying the following conditions:

(i) k is a primary subfield of Ω_m;

---

* See, for example, Gut, loc. cit.

† Hilbert, loc. cit., Theorem 39.

(ii) K is the minimal simple field containing k. We shall call $\{k\}$ the set of fields belonging to K.

By means of Lemma 3, once we have calculated the discriminant of K, we determine at once the discriminant of k, a member of the set of fields belonging to K, if we know merely the relative degree of K/k. Furthermore *if two fields in $\{k\}$ have the same degree their discriminants must coincide.* It is not difficult to determine the conditions K must satisfy in order that there be several fields of the set of equal degree; however, we shall consider only the special case of a K of type $(1, 1, \cdots)$.

Let K be an Abelian field of degree q^n and type $(1, 1, \cdots)$, q an odd prime. From Hilbert's proof of Kronecker's Theorem on Abelian fields, we may deduce that K is a subfield of Ω_m, where

$$m = q^2 q_1 \cdots q_t \quad \text{or} \quad m = q_1 \cdots q_t,$$
$$q_i \equiv 1 \pmod q \qquad (i = 1, \cdots, t),$$

according as q does or does not divide the discriminant of K. Evidently K is simple only if the number of distinct primes dividing m is equal to n, i.e.

$$m = q^2 q_1 \cdots q_{n-1} \quad \text{or} \quad m = q_1 \cdots q_n.$$

Now if K is not simple, it is readily seen that the simple field to which it belongs is itself of type $(1, 1, \cdots)$. Let us then assume K simple, and for the sake of definiteness let us suppose $q \mid m$. Then if the r_i are defined as in (1), K corresponds to the group $(q_0 = q)$

$$G_\mu = \{r_0^q, \cdots, r_{n-1}^q\}.$$

We can now easily determine the set of fields belonging to K:

(i) Let us consider first *all* the cyclic subfields of K; from a well known result concerning Abelian groups, we see at once that the number of such fields is

$$(q^n - 1)/(q - 1) \cdot$$

They may be sorted by considering the number of primes contained in their discriminants. There are first of all n fields whose discriminants contain but a single prime; each corresponds to a subgroup of the type

$$\{r_0^q, r_1, \cdots, r_{n-1}\}.$$

Secondly, there are

$$\binom{n}{2}(q - 1) = \frac{n(n-1)}{2!}(q - 1)$$

fields whose discriminants contain exactly two primes. They fall into $n(n-1)/2$ sets of $(q-1)$ fields, all the fields in a set having the property that their discriminants are divisible by the same primes. Thus a particular set corresponds to

$$\left\{ r_0^q,\ r_0^{a_1} r_1^{-a_0},\ r_1^q,\ r_2,\ \cdots,\ r_{n-1} \right\},$$
$$a_0,\ a_1 = 1, \cdots, q-1, \quad a_0 a_1 \equiv 1 \pmod{q}.$$

Thirdly, there are

$$\binom{n}{3}(q-1)^2$$

fields whose discriminants are divisible by exactly three primes; they fall into $n(n-1)(n-2)/6$ sets of $(q-1)^2$ fields each, all the fields in a set having the property that their discriminants contain the same primes. A particular set corresponds to

$$\left\{ r_0^q,\ r_1^q,\ r_2^q,\ r_0^{a_1} r_1^{-a_0},\ r_0^{a_2} r_2^{-a_0},\ r_3,\ \cdots,\ r_{n-1} \right\},$$
$$a_0,\ a_1,\ a_2 = 1, \cdots, q-1, \quad a_0 a_1 a_2 \equiv 1 \pmod{q}.$$

Finally there are

$$\binom{n}{n}(q-1)^{n-1} = (q-1)^{n-1}$$

fields whose discriminants contain all n primes; they comprise a single set of fields. Each field in the set corresponds to a particular*

$$G^{(a_0, \cdots, a_{n-1})} = \left\{ r_0^q,\ \cdots,\ r_{n-1}^q,\ r_0^{a_1} r_1^{-a_0},\ \cdots,\ r_0^{a_{n-1}} r_{n-1}^{-a_0} \right\}, \tag{7}$$
$$a_i = 1, \cdots, q-1, \quad a_0 a_1 \cdots a_{n-1} \equiv 1 \pmod{q}.$$

It will be convenient for a later application to denote the field corresponding to

$$G^{(a_0, \cdots, a_{n-1})} \text{ by } k^{(a)} = k^{(a_0, \cdots, a_{n-1})}.$$

The fields $k^{(a)}$ are the only primary (cyclic) subfields of K and hence are the only cyclic fields in the set belonging to K.

(ii) To determine $A_n^{(s)}$, the number of fields of degree q^s (and of type $(1, 1, \cdots))$ in the set belonging to K, we notice first that the *total* number of subfields of K of degree q^r (and necessarily of type $(1, 1, \cdots))$ is equal to

* While it may appear from (7) that r_0 plays a special rôle, this is by no means the case. Thus it is easily verified that $G^{(a_0, \cdots, a_{n-1})}$ contains all numbers of the form $r_i^{a_j} r_j^{-a_i}$ and therefore any r_i might be used in place of r_0 in defining the group.

$$\begin{bmatrix} n \\ s \end{bmatrix} = \frac{(q^n - 1) \cdots (q^{n-s+1} - 1)}{(q - 1) \cdots (q^s - 1)}.$$

Then by an argument similar to that employed in the special case (i), we see that

$$\sum_{j=s}^{n} \binom{n}{j} A_j^{(s)} = \begin{bmatrix} n \\ s \end{bmatrix}. \tag{8}$$

To solve (8) for $A_n^{(s)}$ we may proceed thus:

$$\begin{aligned} \sum_{k=s}^{n} (-1)^{n-k} \binom{n}{k} \begin{bmatrix} k \\ s \end{bmatrix} &= \sum_{k=s}^{n} (-1)^{n-k} \binom{n}{k} \sum_{j=s}^{k} \binom{k}{j} A_j^{(s)} \\ &= \sum_{j=s}^{n} (-1)^{n-j} \binom{n}{j} A_j^{(s)} \sum_{k=j}^{n} (-1)^{k-j} \binom{n-j}{k-j} \\ &= \sum_{j=s}^{n} (-1)^{n-j} \binom{n}{j} A_j^{(s)} (1-1)^{n-j} = A_n^{(s)}. \end{aligned} \tag{9}$$

Further transformation of the left member of (9) leads to an unexpected connection between $A_n^{(s)}$ and generalisations of certain important quantities in finite differences. We make use of the formula (the q-generalisation of the binomial theorem)

$$(x + 1)(x + q) \cdots (x + q^{s-1}) = \sum_{\alpha=0}^{s} \begin{bmatrix} s \\ \alpha \end{bmatrix} q^{\alpha(\alpha-1)/2} x^{s-\alpha};$$

then

$$\begin{aligned} (q^n - 1)(q^{n-1} - 1) \cdots (q^{n-s+1} - 1) &= q^{-s(s-1)/2}(q^n - 1)(q^n - q) \cdots (q^n - q^{s-1}) \\ &= q^{-s(s-1)/2} \sum_{\alpha=0}^{s} (-1)^{\alpha} \begin{bmatrix} s \\ \alpha \end{bmatrix} q^{\alpha(\alpha-1)/2 + n(s-\alpha)}, \end{aligned}$$

so that

$$\begin{aligned} A_n^{(s)} &= \frac{q^{-s(s-1)/2}}{(q^s - 1) \cdots (q - 1)} \sum_{k=0}^{n} (-1)^{n-k} \binom{n}{k} \sum_{\alpha=0}^{s} (-1)^{\alpha} \begin{bmatrix} s \\ \alpha \end{bmatrix} q^{\alpha(\alpha-1)/2 + k(s-\alpha)} \\ &= \frac{q^{-s(s-1)/2}}{(q^s - 1) \cdots (q - 1)} \sum_{\alpha=0}^{s} (-1)^{\alpha} \begin{bmatrix} s \\ \alpha \end{bmatrix} (q^{s-\alpha} - 1)^n q^{\alpha(\alpha-1)/2}. \end{aligned} \tag{10}$$

Let us think for the moment of q as an arbitrary parameter, and write $[x]$ for the so-called "basic" number

$$(q^x - 1)/(q - 1)$$

which reduces to x when $q = 1$. Further, defining $[m]!$ by

$$[m]! = [m][m-1]\cdots[1],$$

(10) becomes

$$A_n^{(s)} = (q-1)^{n-s}\frac{q^{-s(s-1)/2}}{[s]!}\sum_{\alpha=0}^{s}(-1)^\alpha\begin{bmatrix} s \\ \alpha \end{bmatrix}[s-\alpha]^n q^{\alpha(\alpha-1)/2}, \tag{11}$$

which, but for the $(q-1)^{n-s}$, is a q-generalisation of what are sometimes called Stirling numbers.

Before leaving the $A_n^{(s)}$ we derive one other important formula connecting them. From (9)

$$\begin{aligned} A_{n+1}^{(s)} - A_n^{(s-1)} &= \sum_{j=s}^{n+1}(-1)^{n+1-j}\left\{\binom{n+1}{j}\begin{bmatrix} j \\ s \end{bmatrix} - \binom{n}{j-1}\begin{bmatrix} j-1 \\ s-1 \end{bmatrix}\right\} \\ &= \sum_{j=s}^{n+1}(-1)^{n+1-j}\left\{\binom{n}{j}\begin{bmatrix} j \\ s \end{bmatrix} + \binom{n}{j-1}\begin{bmatrix} j \\ s \end{bmatrix} \right. \\ &\qquad \left. - \binom{n}{j-1}\begin{bmatrix} j-1 \\ s-1 \end{bmatrix}\right\}; \end{aligned} \tag{12}$$

but

$$\begin{bmatrix} j \\ s \end{bmatrix} - \begin{bmatrix} j-1 \\ s-1 \end{bmatrix} = q^s\begin{bmatrix} j-1 \\ s \end{bmatrix},$$

so that the right member of (12) becomes

$$\begin{aligned} &\sum_{j=s}^{n}(-1)^{n+1-j}\binom{n}{j}\begin{bmatrix} j \\ s \end{bmatrix} + q^s\sum_{j=s+1}^{n+1}(-1)^{n+1-j}\binom{n}{j-1}\begin{bmatrix} j-1 \\ s \end{bmatrix} \\ &\qquad = (q^s-1)\sum_{j=s}^{n}(-1)^{n-j}\binom{n}{j}\begin{bmatrix} j \\ s \end{bmatrix}. \end{aligned}$$

Therefore, finally

$$A_{n+1}^{(s)} = A_n^{(s-1)} + (q^s-1)A_n^{(s)}. \tag{13}$$

5. Simple subfields as relative Abelian fields

We return to the consideration of the general case defined at the beginning of §4, that of a simple subfield K and the set of fields $\{k\}$ belonging to it. By Lemma 2, the relative discriminant of K with respect to any k of $\{k\}$ is the unit ideal of k; further it is clear that K/k is relative Abelian. Let us then for brevity say that K has the property* A with respect to k.

*K is of course part of the Klassenkörper of each k. For definition and proof of the existence of the Klassenkörper of an arbitrary algebraic field, see Furtwängler, Mathematische Annalen, 1907, pp. 1–37; Takagi, Journal of the College of Science, Imperial University of Tokyo, vol. 41 (1920). As no use of the existence of the Klassenkörper is being made here, it is found convenient to use the terminology defined above.

Let $K, k_1, \cdots, k_j$, where the fields $k_1, \cdots, k_j$ are all in $\{k\}$, the set be-belonging to K, be a chain of fields as in §2. Then it is clear that each k has the property A with respect to any succeeding k of the chain. Conversely, we shall now prove that if any Abelian field F have the property A with respect to a k of the chain, then F itself is a member of the chain, and lies somewhere between K and k (possibly at an end).

By hypothesis the relative discriminant of F/k is the unit ideal of k, so that the only primes dividing the discriminant of F are those dividing the discriminant of k and therefore of Ω_m, the cyclotomic field of which k is a primary subfield. Then F is a primary subfield of an Ω_{m}', where

$$m = q_1^{f_1} \cdots q_n^{f_n} \text{ and } m' = q_1^{f_1'} \cdots q_n^{f_n'} \qquad (f_i' \geqq f_i). \tag{14}$$

Let K' be that simple subfield of $\Omega_{m'}$ to which F belongs; clearly K' must have the property A with respect to k. Let μ_i, ν_i, μ, ν be the numbers determining K (see (3) and (4)); $\mu_i', \nu_i', \mu', \nu'$ the corresponding numbers for K'. Let ρ be the relative degree of K/k, w the relative degree of K'/k. Now ν, ν' are the degree of K and K', respectively, so that

$$\nu' = \frac{w\nu}{\rho}. \tag{15}$$

By Lemmas 1 and 3, the discriminants of k and K' are

$$\Pi q_i^{t_i/\rho} \text{ and } \Pi q_i^{t_i'},$$

respectively, where

$$t_i = \frac{\nu}{\phi_i}(s_i - \mu_i + 1), \; t_i' = \frac{\nu'}{\phi_i'}(s_i' - \mu_i' + 1), \tag{16}$$

and s_i, s_i' are defined by (6). If now we use the fact that the relative discriminant of K'/k is the unit ideal,

$$d(K') = d^w(k), \text{ and } t_i' = wt_i/\rho.$$

Using this last equality, together with (15) and (16), we get

$$\frac{s_i - \mu_i + 1}{q_i^{f_i - 1}} = \frac{s_i' - \mu_i' + 1}{q_i^{f_i' - 1}} \qquad (i = 1, \cdots, n),$$

that is,

$$f_i(q_i - 1) - \frac{\mu_i - 1}{q_i^{f_i - 1}} = f_i'(q_i - 1) - \frac{\mu_i' - 1}{q_i^{f_i' - 1}} \quad (i = 1, \cdots, n). \tag{17}$$

Since μ_i and $\mu_i' < q$ it follows from (17), first, that $f_i' = f_i$, and then immediately $\mu_i' = \mu_i$. But this shows that K' and K are identical. We may now state the theorem.

THEOREM 1. *Let K be any simple subfield of Ω_m, and let $\{k\}$ be the set of fields belonging to K. Then K has the property A with respect to each k. Conversely, any Abelian field that has the property A with respect to some k is necessarily a subfield of K.*

Some information about the class number of the fields considered can be derived from this general theorem.* *If F is relative Abelian with respect to G and the relative discriminant of F/G is the unit ideal of G, then the class number of G is divisible by ρ, the relative degree of F/G.* Actually Hilbert proves the theorem only in the case ρ a prime, but as he remarks there is no great difficulty in extending the result to the general case. Hence we obtain

THEOREM 2. *Let k be any primary subfield of Ω_m, and K the minimal simple field containing k. Then if ρ denote the relative degree of K/k, the class number of k is a multiple of ρ.*

6. COMMON INDEX DIVISORS

A rational prime p is called a common index divisor of an arbitrary algebraic field F if, for every integer ω of the field,

$$p \left| \frac{d(\omega)}{d}, \right.$$

where d is the discriminant of F, and $d(\omega)$ that of ω. The following criterion deduced from a result of Dedekind's is given by Hensel.†

Let the prime-ideal decomposition of p in F be

$$p = \mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_r^{e_r}, \quad N(\mathfrak{p}_i) = p^{f_i}. \tag{18}$$

Let $\psi(f)$ denote the number of primary irreducible polynomials (mod p) *of degree f:*

$$\psi(f) = \frac{1}{f} \sum_{d|f} \mu(d) p^{f/d} = \frac{1}{f}\left(p^f - \sum p^{f/p_1} + \sum p^{f/(p_1 p_2)} - \cdots\right). \tag{19}$$

Then a necessary and sufficient condition that p be a common index divisor of F is that, for at least one i,

$$\psi(f_i) < g(f_i),$$

$g(f)$ denoting the number of $\mathfrak{p}$'s in (18) *of degree f.*

---

* Hilbert, loc. cit., Theorem 94.

† Bachmann, *Zahlentheorie* V: *Allgemeine Arithmetik der Zahlenkörper*, 1926, p. 276.

To apply this something must be known about the decomposition of primes in the field to be considered. For an Abelian field this information is given by another theorem of Dedekind's.*

DECOMPOSITION RULE.† *Let Ω_m be a cyclotomic field and F any subfield. Let the group of Ω_m be represented by a reduced residue system* (mod m) *and let* (h) *denote the subgroup corresponding to F. Let p^k be the highest power of the prime p dividing m, $m=p^k m'$; and let the number of those numbers of (h) that are $\equiv 1$(mod m') be $\phi(p^k)/g$, thus defining g. Let f be the smallest positive integer such that*

$$p^f \equiv (h) \pmod{m'}, \tag{20}$$

that is, to one of the numbers in (h). Then the prime-ideal decomposition of p in F is

$$p = (\mathfrak{p}_1 \cdots \mathfrak{p}_e)^g, \quad N(\mathfrak{p}_i) = p^f,$$

where $e\cdot f\cdot g$ is the degree of F.

We take first the simplest and perhaps the most interesting case, that of a cyclic field C of odd prime degree q and of discriminant divisible by a single prime. Then the discriminant is, by Kronecker's Theorem and Lemma 1, either

$$\text{(a)} \quad q^{2(q-1)}; \quad \text{or (b)} \quad l^{q-1},$$

where l is a prime such that $l\equiv 1$ (mod q). By the Decomposition Rule (or directly, using well known theorems on the decomposition of a prime in a Galois field) the condition that a prime p factor in C is either

$$(21) \text{ or } \begin{array}{lll} \text{(a)} & p^{q-1} \equiv 1 \pmod{q^2} & (p \neq q); \\ \text{(b)} & p^{(l-1)/q} \equiv 1 \pmod{l} & (p \neq l). \end{array}$$

Now if p factor in C it factors into q distinct prime ideals (of the first degree). Hence, applying the criterion for common index divisors, and noticing that $\psi(1)=p$, we deduce one of the theorems stated in the Introduction.

THEOREM 3.‡ *Let C be a cyclic field of prime degree q and of discriminant divisible by a single prime. Then a necessary and sufficient condition that a prime $p(p<q)$ be a common index divisor of C is furnished by equations* (21).

* Gesammelte Werke, vol. 1, p. 233.

† This theorem is implicitly proved by Gut, loc. cit., §§5 and 8.

‡ For an equivalent criterion for cubic fields see Hensel, Journal für Mathematik, vol. 113 (1894), p. 147.

The necessity of the condition follows from the theorem* that a common index divisor of any field is less than the degree of the field.

Turning now to the case of the general cyclic field C of odd prime degree, we remark first that its discriminant, d, is either

$$\text{(a)}\quad (q^2q_1\cdots q_n)^{q-1};\quad\text{or}\quad\text{(b)}\quad (q_1\cdots q_n)^{q-1};\quad q_i\equiv 1 \pmod{q}.$$

Using the notation of §4 (i), let C correspond to the group

$$\text{(a)}\qquad G^{(a_0,\cdots,a_n)}\quad\text{or (b)}\qquad G^{(a_1,\cdots,a_n)}.$$

To determine the condition that a prime p(p does not divide d) factor in C we use the Decomposition Rule. We need consider but one case in detail; let us take case (a). It is plain that $m=q^2q_1\cdots q_n$, and clearly the condition that p factor is that f in (20) be one; or putting

$$p\equiv r_0^{c_0}\cdots r_n^{c_n}\pmod{m}\qquad (1\leqq c_i\leqq \phi(q_i)) \tag{22a}$$

p factors provided that integers s, t can be found such that

$$r_0^{c_0}\cdots r_n^{c_n}\equiv r_0^{qs_0}\cdots r_n^{qs_n}(r_0^{a_1}r_1^{-a_0})^{t_1}\cdots(r_0^{a_n}r_n^{-a_0})^{t_n}\pmod{m}.$$

But this congruence is equivalent to the system

$$c_0\equiv qs_0+a_1t_1+\cdots+a_nt_n\pmod{\phi(q^2)},$$
$$c_i\equiv qs_i-a_0t_i\pmod{\phi(q_i)}\qquad(i=1,\cdots,n),$$

which is equivalent to

$$a_0c_0+\cdots+a_nc_n\equiv 0\pmod{q}, \tag{23a}$$

the condition sought.

THEOREM 4. *A necessary and sufficient condition that p ($p<q$) be a common index divisor of $C=C^{(a_0,\cdots,a_n)}$ is furnished by* (22a) *and* (23a). *Similarly a necessary and sufficient condition that p($p<q$) be a common index divisor of*

$$C=C^{(a_1,\cdots,a_n)}$$

is furnished by

$$p\equiv r_1^{c_1}\cdots r_n^{c_n}\pmod{m}, \tag{22b}$$

and

$$a_1c_1+\cdots+a_nc_n\equiv 0\pmod{q}. \tag{23b}$$

---

* Proved by von Zylinski, Mathematische Annalen, vol. 73 (1913), p. 273.

Turning next to the simple field K defined by (3) and (4), the Decomposition Rule shows that if p does not divide $d(K)$,

$$p = \mathfrak{p}_1 \cdots \mathfrak{p}_e, \quad N(\mathfrak{p}_i) = f, \quad ef = \nu;$$

and

$$f \mid \omega, \quad \omega = \text{L.C.M.}(\nu_1, \cdots, \nu_n).$$

If then $\psi(f)$, the number of primary irreducible polynomials (mod p) of degree f, is less than e, p is a common index divisor of K. But evidently

$$\psi(f) \leqq p^f \leqq p^\omega;$$

and

$$e = \frac{\nu}{f} \geqq \frac{\nu}{\omega} \cdot$$

If then $\omega p^\omega < \nu$, surely $\psi(f) < e$. Hence we have

THEOREM 5. *Let K be the simple field of degree ν defined by* (3) *and* (4). *If p does not divide $d(K)$, and*

$$\omega p^\omega < \nu, \quad \omega = \text{L.C.M.}(\nu_1, \cdots, \nu_n), \tag{24}$$

then p is surely a common index divisor of K. The inequality (24) *may be replaced by the weaker condition*

$$\omega \cdot \operatorname*{Max}_{f \mid \omega} \psi(f) < \nu. \tag{24$'$}$$

Theorem 5 could without much difficulty be refined in several directions. And it would also be possible to frame a great many theorems analogous to Theorem 4 for various kinds of Abelian fields. However we shall limit ourselves to the case of fields of type $(1, 1, \cdots)$. Assume first that the prime p does not divide the discriminant of the field. Then by the Decomposition Rule or directly it may easily be shown that either

$$p = \mathfrak{p}_1 \cdots \mathfrak{p}_{q^n}, \text{ each } \mathfrak{p} \text{ of degree } 1; \tag{i}$$

or

$$p = \mathfrak{p}_1 \cdots \mathfrak{p}_{q^{n-1}}, \text{ each } \mathfrak{p} \text{ of degree } q; \tag{ii}$$

the field being of degree q^n. If p divides the discriminant, we get, in place of (i) and (ii),

$$p = (\mathfrak{p}_1 \cdots \mathfrak{p}_{q^{n-1}})^q, \text{ each } \mathfrak{p} \text{ of degree } 1; \tag{iii}$$

or,

(iv) $$p = (\mathfrak{p}_1 \cdots \mathfrak{p}_{q^{n-2}})^q, \text{ each } \mathfrak{p} \text{ of degree } q.$$

Now

$$\psi(1) = p, \text{ and } \psi(q) = \frac{p^q - p}{q}.$$

Application of the Hensel criterion leads to

THEOREM 6. *Let K be of degree q^n and type $(1, 1, \cdots)$. Then if*

(i) $$p \textit{ does not divide } d(K), \quad p \leqq q^{n/q};$$

or

(ii) $$p \mid d(K), \qquad p \leqq q^{(n-1)/q},$$

p is surely a common index divisor of K.

It is perhaps worth remarking that in Theorem 6 either q or $d(K)$ may be even.

Theorem 6 evidently implies that, if q be fixed, then, for sufficiently large n, an assigned prime p will be a common index divisor in any field of type $(1, 1, \cdots$ to n units). Thus for example the primes 2, 3, 5, 7 are common index divisors of

$$k((-3)^{1/2}, 5^{1/2}, (-11)^{1/2}, 13^{1/2}, 17^{1/2}, (-19)^{1/2}).$$

We consider finally a refined form of Theorem 6 for the case in which the (odd) discriminant is divisible by exactly n primes. The field is then simple. To determine the decomposition of rational primes in such a field we could of course apply once more the decomposition rule. It is however somewhat simpler and perhaps more interesting to proceed differently. The field K under consideration is, by Kronecker's Theorem, composed of the n cyclic fields $C(q_i)$, each of degree q and of discriminant a power of q_i. Here q_i is either a prime $\equiv 1 \pmod q$; or, if $q \mid d(K)$, one of them is q^2. From Theorem 3 we already know when a prime $p (p \neq q_i)$ will factor in $C(q_i)$; as for q_i we have of course (in $C(q_i)$) either

$$q_i = \mathfrak{q}^q,$$

or

$$q = \mathfrak{q}^q \text{ for } q_i = q^2.$$

Now p may decompose in K in one of four possible ways (see the proof of the preceding theorem). It is now fairly clear that if p does not divide $d(K)$, and

(25) $$p^{\phi(q_i)/q} \equiv 1 \pmod{q_i} \quad \text{for} \quad i = 1, \cdots, n,$$

then

$$p = \mathfrak{p}_1 \cdots \mathfrak{p}_{q^n}, \; N(\mathfrak{p}_i) = p;$$

if (25) fails for at least one i, then

$$p = \mathfrak{p}_1 \cdots \mathfrak{p}_{q^{n-1}}, \; N(\mathfrak{p}_i) = p^q;$$

if $p \,|\, d(K)$, and

$$(25)' \qquad p^{\phi(q_i)/q} \equiv 1 \pmod{q_i},$$

for all i such that p does not divide q_i, then

$$p = (\mathfrak{p}_1 \cdots \mathfrak{p}_{q^{n-1}})^q, \; N(\mathfrak{p}_i) = p;$$

but if (25)′ fail for at least one i, then

$$p = (\mathfrak{p}_1 \cdots \mathfrak{p}_{q^{n-2}})^q, \; N(\mathfrak{p}_i) = p^q.$$

We are now able to apply the Hensel criterion and we have at once

THEOREM 7. *Let K be of degree q^n and type* $(1, 1, \cdots)$; *and let $d(K)$ be divisible by exactly n primes. Let p be a prime $<q^n$; then if p does not divide $d(K)$, and*

(i) *if* (25) *hold, p is a common index divisor;*

(ii) *if* (25) *fails for at least one i, then p is a common index divisor only if*

$$p^q - p < q^n;$$

if $p \,|\, d(K)$, and

(iii) *if* (25)′ *hold, p is a common index divisor if*

$$p < q^{n-1};$$

(iv) *if* (25)′ *fails for at least one i, then p is a common index divisor only if*

$$p^q - p < q^{n-1}.$$

CAMBRIDGE UNIVERSITY,
CAMBRIDGE, ENGLAND

ON THE DERIVATIVES OF NEWTONIAN AND LOGARITHMIC POTENTIALS NEAR THE ACTING MASSES*

BY

MILDRED M. SULLIVAN

1. Introduction. The existing theorems† on the continuity of the derivatives of the potentials of various spreads of acting matter, at points of the spreads, exact more of the densities and surfaces involved than is necessary for the conclusions drawn. A single exception is the work of Petrini,‡ who considers necessary and sufficient conditions. The generality of his results, and the incident delicacy of the considerations establishing them, have prevented their becoming widely current. Moreover, they are concerned only with the derivatives of the first two orders in the case of potentials of volume distributions, and of the first order in the case of surface spreads.

Recently,§ Professor Kellogg, in a study of the continuity of harmonic functions defined by their boundary values, and of their derivatives, has shown the usefulness of a simple condition due to Dini.|| The present paper consists in a systematic application of the Dini condition to harmonic functions defined by the densities of the spreads whose potentials they are. Because of the immediate availability of the same methods of proof, theorems have been added on the effect of Hölder conditions on the densities and their derivatives in assuring the existence of Hölder conditions on the potentials and their derivatives, results which are already at hand only for the derivatives of the first order, in the work of Schauder (loc. cit.). The treatment is

---

* Presented to the Society, October 31, 1931; received by the editors April 20, 1932.

The author desires to make to Professor O. D. Kellogg grateful acknowledgment for the suggestion which led to this paper, and for his helpful interest in its development.

† See, for the earlier literature, L. Lichtenstein, *Neuere Entwickelung der Potentialtheorie. Konforme Abbildung*, Encyklopädie der mathematischen Wissenschaften, II C 3, pp. 199–209. See also J. Schauder, *Potentialtheoretische Untersuchungen*, Erste Abhandlung, Mathematische Zeitschrift, vol. 33 (1931), pp. 602–640; L. Lichtenstein, *Über einige Hilfssätze der Potentialtheorie*, IV, Sächsische Berichte, vol. 82 (1930), pp. 265–344.

‡ *Les dérivées premières et secondes du potentiel*, Acta Mathematica, vol. 31 (1908), pp. 127–332.

§ *On the derivatives of harmonic functions on the boundary*, these Transactions, vol. 33 (1931), pp. 486–510.

|| *Sur la méthode des approximations successives pour les équations aux dérivées partielles du deuxième ordre*, Acta Mathematica, vol. 25 (1902), p. 224. The form of the condition used in the present paper is slightly more general than that employed by Professor Kellogg.

given for space of three dimensions, but it is shown that the results are valid also for logarithmic potentials.*

2. **Definitions and lemmas.** We introduce the following

Definition. Let $f(q)$ be defined on the regular surface element S†, and be continuous at the point p of S. Let q vary on S in any fixed plane through the normal to S at p. Then $f(q)$ is said to satisfy *a uniform Dini condition at p if the integral*

$$\int_0^a \frac{|f(q) - f(p)|}{r} dr, \quad r = \overline{pq}, \ddagger \tag{1}$$

converges uniformly as to the normal plane chosen. $f(q)$ is said to satisfy *a uniform Dini condition on S if it is continuous on S and the convergence of the integral* (1) *is uniform both as to the point p and as to the direction at p.*

It will be convenient to use the following class notations. If S is a regular surface element admitting for some orientation of the axes a representation $\zeta = \phi(\xi, \eta)$ where $\phi(\xi, \eta)$ has derivatives of order n which satisfy a uniform Dini condition on S we shall say that S *is a regular surface element of class* $C^{n+\delta}$. If $f(\xi, \eta)$ has derivatives of order n on S which satisfy a uniform Dini condition there, we shall say that $f(\xi, \eta)$ is *of class* $C^{n+\delta}$ *on* S. To say that S (or $f(\xi, \eta)$) *is of class* $C^{n+\lambda}$ will mean that the derivatives of $\phi(\xi, \eta)$ (or $f(\xi, \eta)$) of order n satisfy a uniform Hölder condition with exponent λ on S.

A *region of type V* will be understood to mean a closed region of space (consisting of an open continuum and its boundary), partially bounded by the surface S under consideration, but containing no boundary points of S, and such that the point P can approach S from only one side while remaining in V.

We shall frequently have to consider the product or quotient of two functions each of which satisfies a uniform Dini condition on S. We shall need the following lemmas.

Lemma 1. *Let f and ϕ be continuous functions satisfying a uniform Dini condition on S. Then $f\phi$ satisfies a uniform Dini condition on S, and, provided ϕ has a positive lower bound, so also does f/ϕ.*

This follows at once from the inequalities

---

* In a paper entitled *Potential functions on the boundary of their regions of definition*, these Transactions, vol. 9 (1908), pp. 39–50, Professor Kellogg used a condition of the Dini type in a study of double distributions in the plane. The present results are more general in that less is required of the boundary curve.

† For definition, see Kellogg, *Foundations of Potential Theory*, Berlin, 1929, p. 105.

‡ It is readily shown on the basis of the regularity of S that, for sufficiently small a, to each value of r, $0 \leqq r \leqq a$, there corresponds one and only one point q.

$$\int_\epsilon^a \frac{|f(q)\phi(q)-f(p)\phi(p)|}{r}dr$$
$$\leqq \int_\epsilon^a \frac{|\phi(q)||f(q)-f(p)|}{r}dr + \int_\epsilon^a \frac{|f(p)||\phi(q)-\phi(p)|}{r}dr$$
$$\leqq \max|\phi| \int_\epsilon^a \frac{|f(q)-f(p)|}{r}dr + \max|f| \int_\epsilon^a \frac{|\phi(q)-\phi(p)|}{r}dr$$

together with similar ones for the quotient.

In the same way, we establish

LEMMA 2. *If S is of class $C^{1+\delta}$,* $\sec\gamma = (1+\phi_\xi^2+\phi_\eta^2)^{1/2}$ *satisfies a uniform Dini condition on S.*

Here γ denotes the angle between the normal to S at q and the ζ-axis. As a corollary to these lemmas, we note that if S is in $C^{1+\delta}$ and if $f(q)$ satisfies a uniform Dini condition on S, the same is true of $f(q)\sec\gamma$.

We shall also have need of

LEMMA 3. *Let S be a regular surface element, and let r' denote the projection of $r=\overline{pq}$ on the tangent plane to S at p. Then if either of the integrals*

$$\int_0^a \frac{|f(q)-f(p)|\,dr'}{r'}, \quad \int_0^a \frac{|f(q)-f(p)|}{r}dr$$

converges uniformly as to any parameters, so does the other.

This follows from the facts, first that $r'\leqq r$, and secondly that $r\leqq$ arc $\overset{\frown}{pq} < r' \max|\sec\gamma|$, the arc being in the plane through p and q containing the normal to S at p. The existence of the maximum of $\sec\gamma$ is implied in the definition of regular surface element. Thus the above integrals may be used interchangeably in Dini conditions.

It is important to notice the following.

LEMMA 4. *The definition of the classes $C^{n+\delta}$ and $C^{n+\lambda}$ is independent of the coördinate system. That is, if S is in either class for one system of orthogonal axes, it is in the same class for any other system of orthogonal axes in which the angle between the tangent plane to the surface and the ζ-axis has a positive lower bound.*

If x, y, z are the coördinates referred to the given axes and ξ, η, ζ the coördinates referred to any other axes with the specified properties, we have on S

$$\xi = l_1x + m_1y + n_1f(x,y), \quad \eta = l_2x + m_2y + n_2f(x,y),$$
$$\zeta = l_3x + m_3y + n_3f(x,y),$$

where $z=f(x, y)$ is the given representation of S. These equations define a representation $\zeta=\phi(\xi, \eta)$ of S with respect to the second system of axes. We find, then,

$$\frac{\partial\zeta}{\partial\xi} = -\frac{l_1f_x + m_1f_y - n_1}{l_3f_x + m_3f_y - n_3}.$$

The denominator is the cosine of the angle between the normal to S and the ζ-axis multiplied by $(1+f_x^2+f_y^2)^{1/2}$. It has, therefore, a positive lower bound. Then, if S is of class $C^{1+\delta}$ in the given system of coördinates, it follows that this derivative satisfies a uniform Dini condition on S. The same considerations apply to the other derivative of the first order of ζ. Therefore, S is of class $C^{1+\delta}$ in the second system of coördinates. Similarly, if S is of class $C^{1+\lambda}$ in the given system of coördinates, the derivatives of the first order of ζ satisfy a uniform Hölder condition on S.

In the same way, Lemma 4 can be extended to derivatives of any order. Thus the concept of the class of a regular surface element S is independent of the coördinate system provided no normal to the surface is at right angles to the ζ-axis.

If S is a regular surface element there exists a positive constant c such that if σ is the portion of S in a sphere of radius c about any interior point of S, then σ has a standard representation with tangent-normal axes at any point p of σ.* Applying Lemma 4, we see that if S is of class $C^{n+\delta}$, σ is of class $C^{n+\delta}$ with respect to a tangent-normal system of axes at any point p of σ.

In our theorems on the derivatives of the first order of the potential of a simple distribution we need the following.

LEMMA 5. *If S is a regular surface element of class $C^{1+\delta}$ having a standard representation with respect to a tangent-normal system of axes at any interior point p given by $\zeta=\phi(\xi, \eta)$, the integral*

$$\iint_{\sigma'} \frac{|\zeta|}{r'^3}\,dS',$$

where σ' is a circle about p in the (ξ, η)-plane and r' is the projection of r on this plane, *converges at p, and the convergence is uniform as to p.*

We cut out from σ' a circle c of radius ϵ. Then

$$\iint_{\sigma'-c} \frac{|\zeta|}{r'^3}\,dS' = \int_0^{2\pi}\int_\epsilon^a \frac{|\zeta|}{r'^2}\,dr'd\theta,$$

* See Kellogg, *Foundations of Potential Theory*, loc. cit., p. 157.

since the integrand is continuous in $\sigma'-c$. This equality holds for any $\epsilon>0$. If the inner integral on the right converges uniformly as to θ when $\epsilon\to 0$, the iterated integral approaches a limit, which limit is the value of the improper double integral over σ', as well as the value of the iterated integral in which r' goes from 0 to a.

We shall study, then,

$$\int_\epsilon^a \frac{|\zeta|}{r'^2}\,dr'.$$

If we write $\xi = t\cos\theta$, $\eta = t\sin\theta$,

$$\begin{aligned}\zeta &= \zeta(0,0) + \int_0^{r'} \frac{d\zeta}{dt}\,dt \\ &= 0 + \int_0^{r'} [\phi_\xi(t\cos\theta, t\sin\theta)\cos\theta + \phi_\eta(t\cos\theta, t\sin\theta)\sin\theta]dt.\end{aligned}$$

Hence,

$$\begin{aligned}|\zeta| &\leqq \int_0^{r'} \{\,|\phi_\xi(t)| + |\phi_\eta(t)|\,\}dt \\ &= \int_0^{r'} t\frac{d}{dt}(I+J)dt,\end{aligned}$$

where

$$I(t,\theta) = \int_0^t \frac{|\phi_\xi(r'\cos\theta, r'\sin\theta)|}{r'}\,dr',\ J(t,\theta) = \int_0^t \frac{|\phi_\eta(r'\cos\theta, r'\sin\theta)|}{r'}\,dr'.$$

By hypothesis, I and J vanish with t uniformly as to θ. Therefore,

$$\int_\epsilon^a \frac{|\zeta|}{r'^2}\,dr' \leqq \int_\epsilon^a \frac{1}{r'^2}\int_0^{r'} t\frac{d}{dt}(I+J)dtdr'.$$

Integrating by parts, we have for the integral on the right

$$\begin{aligned}&-\frac{1}{r'}\int_0^{r'} t\frac{d}{dt}(I+J)dt\Big|_\epsilon^a + \int_\epsilon^a \frac{1}{r'}\,r'\,\frac{d}{dr'}(I+J)dr' \\ &= \frac{1}{\epsilon}\int_0^\epsilon t\frac{d}{dt}(I+J)dt - \frac{1}{a}\int_0^a t\frac{d}{dt}(I+J)dt + (I+J)_{r'=a} - (I+J)_{r'=\epsilon}.\end{aligned}$$

The last term vanishes with ϵ. This is true of the first term, also; since, by the law of the mean,

$$\frac{1}{\epsilon}\int_0^\epsilon t\frac{d}{dt}(I+J)dt = \frac{\bar{t}}{\epsilon}\int_0^\epsilon \frac{d}{dt}(I+J)dt$$
$$= \frac{\bar{t}}{\epsilon}(I+J)_{t=\epsilon}$$

where $0<\bar{t}/\epsilon<1$. Therefore,

$$\int_0^a \frac{|\zeta|}{r'^2}dr' \leqq (I+J)_{r'=a} - \frac{1}{a}\int_0^a t\frac{d}{dt}(I+J)dt$$
$$\leqq (I+J)_{r'=a},$$

since the derivatives of I and J with respect to t are never negative. It follows therefore, not only that

$$\iint_{\sigma'} \frac{|\zeta|}{r'^3}dr'$$

converges everywhere on S, but that this integral approaches 0 with the radius of σ', uniformly.

3. **Existence and continuity of the derivatives of the first order of the potential of a simple distribution.**

Tangential derivatives. We prove the following theorem:

THEOREM I. *Let S be a regular surface element of class $C^{1+\delta}$. Let σ, the density of a simple distribution on S, satisfy a uniform Dini condition at the interior point p of S. Then the derivative of the potential U at P in the direction of any tangent at p approaches a limit as P approaches p along the normal. If σ satisfies a uniform Dini condition over a closed portion of S containing no boundary points of S, the limits of such derivatives are approached uniformly as to p on such a portion.*

We shall restrict ourselves to the portion of S contained in a sphere of radius c about p, such that we have a single representation of the whole piece with a tangent-normal system of axes at any point of the piece. We have seen that such a positive constant c exists uniformly as to p. As the potential of the rest of S is analytic in a neighborhood of p, we may neglect it and assume that all of S is contained in this sphere.

We take a tangent-normal system of axes at p, choosing the x-axis in the direction in which we are taking the derivative and the y-axis in the perpendicular tangential direction. Let P be a point on the z-axis. Then for $z\neq 0$,

$$\frac{\partial U}{\partial x} = \iint_S \sigma\frac{\xi}{r^3}dS = \iint_{S'} \sigma \sec\gamma \frac{\xi}{r^3}dS',$$

S' being the projection of S on the (x, y)-plane and γ the angle between the normal to S and the z-axis. This derivative may be written

$$\frac{\partial U}{\partial x} = I + J$$

where

$$I = \iint_{\sigma'} \sigma \sec\gamma \frac{\xi}{r^3} dS', \quad J = \iint_{S'-\sigma'} \sigma \sec\gamma \frac{\xi}{r^3} dS',$$
$$r^2 = \xi^2 + \eta^2 + (z - \zeta)^2,$$

σ' being a small circle of radius a about p in the (x, y)-plane. Then, for any fixed σ', J is continuous, and corresponding to any $\epsilon > 0$, there is a δ depending on ϵ and σ' only, such that

$$| J(z_2) - J(z_1) | < \epsilon/3 \text{ when } 0 < z_1 < \delta, \ 0 < z_2 < \delta.$$

We write

$$I = I_1 + I_2,$$

where

$$I_1 = \sigma(p) \iint_{\sigma'} \frac{\xi}{r^3} dS', \quad I_2 = \iint_{\sigma'} [\sigma(q) \sec\gamma - \sigma(p)] \frac{\xi}{r^3} dS',$$

and compare I_1 with the integral

$$\sigma(p) \iint_{\sigma'} \frac{\xi}{\rho^3} dS' \text{ where } \rho^2 = \xi^2 + \eta^2 + z^2.$$

This integral vanishes since the integrand has equal and opposite values at (ξ, η) and $(-\xi, \eta)$. Hence,

$$\begin{aligned} I_1 &= \sigma(p) \iint_{\sigma'} \xi \left(\frac{1}{r^3} - \frac{1}{\rho^3} \right) dS' \\ &= \sigma(p) \iint_{\sigma'} \frac{\xi\zeta(2z - \zeta)}{r\rho(r + \rho)} \left(\frac{1}{r^2} + \frac{1}{r\rho} + \frac{1}{\rho^2} \right) dS'. \end{aligned}$$

Since

$$| \xi | \leqq r' = (\xi^2 + \eta^2)^{1/2}, | z | \leqq \rho, \ | z - \zeta | \leqq r, | 2z - \zeta | \leqq r + \rho, r' \leqq \rho, r' \leqq r,$$

it follows that

$$| I_1 | \leqq 3 \max | \sigma | \iint_{\sigma'} \frac{| \zeta |}{r'^3} dS'.$$

By Lemma 5, the last integral converges. It follows that we can so choose σ' that

$$\iint_{\sigma'} \frac{|\zeta|}{r'^3}\, dS' < \frac{1}{18 \max |\sigma|}\epsilon,$$

unless $\max|\sigma|=0$ (in which case $I_1 \equiv 0$). Therefore, by choosing a small enough we can make $|I_1| < \epsilon/6$, independently of z. Then

$$|I_1(z_2) - I_1(z_1)| < \epsilon/3.$$

As to the second integral, we have

$$\begin{aligned} |I_2| &\leqq \iint_{\sigma'} |\sigma(q)\sec\gamma - \sigma(p)| \frac{1}{r'^2} dS' \\ &\leqq \int_0^{2\pi}\int_0^a \frac{|\sigma(q)\sec\gamma - \sigma(p)|}{r'} dr' d\theta. \end{aligned}$$

By Lemmas 1 and 2, the inner integral converges at p, uniformly as to θ. Therefore, by taking a small enough, a depending on p in this case, we can make

$$\int_0^a \frac{|\sigma(q)\sec\gamma - \sigma(p)|}{r'} dr' < \epsilon/(12\pi),$$

independently of θ. Then we shall have

$$|I_2(z_2) - I_2(z_1)| < \epsilon/3.$$

Having chosen a we can, as we have seen, by taking $0 < z_1 < \delta$, $0 < z_2 < \delta$, make

$$|J(z_2) - J(z_1)| < \epsilon/3.$$

Therefore, for z_1 and z_2 so restricted,

$$\left| \frac{\partial U(z_2)}{\partial x} - \frac{\partial U(z_1)}{\partial x} \right| < \epsilon,$$

and the derivative approaches a limit as P approaches p along the normal at p.

The only inequality which depends on the point p is that for I_2. If σ satisfies a uniform Dini condition on a closed portion of S this inequality can be made independent of the position of p on this portion of S. Therefore, under these conditions, the approach of the derivative to its limit along the normal will be uniform on this part of S, both as to the point p, and as to the tangential direction of differentiation.

Normal derivatives. In studying the normal derivative at P as P approaches p along the normal, we need assume only that σ is bounded and integrable on S and continuous at p. We assume that S, as in the case of tangential derivatives, is a regular surface element of class $C^{1+\delta}$.

When $z\neq 0$ the normal derivative at $P(0, 0, z)$ is given by

$$\frac{\partial U}{\partial z} = \iint_S \sigma \frac{\zeta - z}{r^3} dS = \iint_{S'} \sigma \sec\gamma \frac{\zeta - z}{r^3} dS' \tag{2}$$

where S' is the projection of S on the (x, y)-plane. Let U' be the potential of a plane lamina occupying the area S' and having the density $\sigma \sec \gamma$. Then

$$\frac{\partial U'}{\partial z} = - \iint_{S'} \sigma \sec\gamma \frac{z}{\rho^3} dS' \quad \text{where } \rho^2 = \xi^2 + \eta^2 + z^2. \tag{3}$$

This derivative, as is well known, approaches $\mp 2\pi\sigma(p)$ according as P approaches p along the positive or the negative z-axis.

We consider the difference

$$\begin{aligned} \frac{\partial U}{\partial z} - \frac{\partial U'}{\partial z} &= \iint_{S'} \sigma \sec\gamma \left[\frac{\zeta - z}{r^3} + \frac{z}{\rho^3}\right] dS' \\ &= \iint_{S'} \sigma \sec\gamma \left[\frac{\zeta}{r^3} - z\left(\frac{1}{r^3} - \frac{1}{\rho^3}\right)\right] dS' \\ &= I_1 + I_2 \end{aligned} \tag{4}$$

where

$$I_1 = \iint_{\sigma'} \sigma \sec\gamma \left[\frac{\zeta}{r^3} - z\left(\frac{1}{r^3} - \frac{1}{\rho^3}\right)\right] dS',$$

$$I_2 = \iint_{S'-\sigma'} \sigma \sec\gamma \left[\frac{\zeta}{r^3} - z\left(\frac{1}{r^3} - \frac{1}{\rho^3}\right)\right] dS'.$$

The integral I_2 is continuous in z for fixed σ'. As to I_1, if we use the algebraic identity already employed in connection with the tangential derivatives, and the inequalities

$$|z| \leqq \rho, \quad |2z - \zeta| \leqq r + \rho, \quad r' \leqq r, \quad r' \leqq \rho,$$

we find that

$$|I_1| \leqq 4 \max|\sigma \sec\gamma| \iint_{\sigma'} \frac{|\zeta|}{r'^3} dS'.$$

Thus, by Lemma 5, I_1 can be made arbitrarily small by sufficiently restricting σ'.

We conclude that the difference (4) is continuous at $z=0$. The value of this difference at p is given by the convergent integrals

$$\iint_{S'} \sigma \sec\gamma \frac{\zeta}{r^3}\, dS' = \iint_S \sigma \frac{\zeta}{r^3}\, dS = \iint_S \sigma \left(\frac{\partial}{\partial z}\frac{1}{r}\right)_p dS.$$

Using the values of the limits of the derivative (3), we have as limits of the derivative (2) when P approaches p from the positive or negative side

$$\left(\frac{\partial U}{\partial z}\right)_+ + 2\pi\sigma(p) = \left(\frac{\partial U}{\partial z}\right)_- - 2\pi\sigma(p) = \iint_S \sigma \left(\frac{\partial}{\partial z}\frac{1}{r}\right)_p dS.$$

These limits are approached uniformly as to p on any closed interior portion of S on which σ is continuous.

Letting n denote the direction of the normal in the positive sense to S at p, we may state the results obtained, in

THEOREM II. *Let S be a regular surface element of class $C^{1+\delta}$. Let σ, the density of a simple distribution on S, be continuous at p. Then the normal derivative of the potential of this distribution approaches a limit as P approaches p along the normal to S at p from either side and these limits are*

$$\frac{\partial U}{\partial n_+} = -2\pi\sigma(p) + \iint_S \sigma\left(\frac{\partial}{\partial n}\frac{1}{r}\right)_p dS,$$

$$\frac{\partial U}{\partial n_-} = +2\pi\sigma(p) + \iint_S \sigma\left(\frac{\partial}{\partial n}\frac{1}{r}\right)_p dS.$$

These limits are approached uniformly as to p on any closed portion of S, containing no boundary points of S, on which σ is continuous.

Derivatives in any direction. It follows from Theorems I and II, that if S is a regular surface element of class $C^{1+\delta}$ and σ satisfies a uniform Dini condition at p, the derivative in any direction approaches a limit as P approaches p along the normal at p. If σ satisfies a uniform Dini condition on a closed interior portion of S, the derivative of U in a fixed direction approaches its limits uniformly along the normals on this portion of S. We shall now prove

THEOREM III. *Let S be a regular surface element of class $C^{1+\delta}$. Let σ satisfy a uniform Dini condition on S. Then the potential U of the distribution of density σ on S has derivatives of the first order which, when defined on S by their limits, are continuous in any region of type V.*

The difficulty of the situation arises from the fact that, in the absence of hypotheses assuring curvatures of S, we cannot count on the existence of a field of normals to S. The following lemma, however, will be all that is needed:

Let Q be a point on the normal to S at p, and let σ denote a sphere about Q. Then there exists a sphere σ' about p, such that every point P of σ' lies on a normal to S which meets σ.

Let α denote the radius of a sphere σ' about p. We shall show that α can be chosen small enough to meet the requirements of the lemma. As a first restriction, 2α is to be less than the distance from the interior point p to the boundary of S. Let P be a point of σ'. The largest sphere about P which contains in its interior no points of S, has on its surface at least one interior point of S. Let q be such a point. Then Pq, being a minimal segment from P to S, is normal to S at q. Moreover, $\overline{Pq} \leqq \overline{Pp} \leqq \alpha$, and hence $\overline{pq} \leqq 2\alpha$. Furthermore, because of the continuity of the direction cosines of the normal to S, the normals to S at points in σ' make with the normal at p, angles having a maximum η, which approaches 0 with α. It follows that the greatest distance from Q to the normal Pq does not exceed $2\alpha + \overline{pQ} \sin \eta$. As $\eta \to 0$ with α, α can be chosen positive and so small that the normal Pq will certainly meet σ.

Turning now to the proof of the theorem, we consider the continuity of any derivative of U at an interior point p of S. Let $F(P)$ denote the derivative in question. $F(P)$ approaches its limits along normals, uniformly. There thus corresponds to any $\epsilon > 0$, a one-sided neighborhood N of p, in which $F(P)$ differs from its limiting values on normals by less than $\epsilon/4$. Let Q be a point of the normal at p in N. Then since $F(P)$ is continuous at Q, there is a sphere σ about Q and in N, such that if P' is in σ,

$$| F(P') - F(Q) | < \epsilon/4.$$

If P is any point in the sphere σ', which corresponds, by the lemma, to σ, we may take P' as a point of the normal Pq. We then have the following inequalities, resulting from the uniform approach along the normals:

$$| F(P') - F(q) | < \epsilon/4,$$
$$| F(P) - F(q) | < \epsilon/4,$$
$$| F(Q) - F(p) | < \epsilon/4.$$

It follows from these that if P is any point in σ',

$$| F(P) - F(p) | < \epsilon,$$

and the derivative is thus continuous at p. The rest of the theorem follows from the analytic character of the potential at points not on S.

4. **Existence and continuity of the potential of a double distribution at points of the distribution.** We shall consider the potential of a double distribution on a regular surface element S of class $C^{1+\delta}$. The moment μ is to be bounded and integrable. The potential is given by the integral

$$U = \iint_S \mu \frac{\partial}{\partial \nu} \frac{1}{r} dS.$$

For the sake of completeness, we show first that this improper integral converges on S under the present assumptions. Using as the field of integration the projection of S on the tangent plane at p, this amounts to verifying that

$$\lim \iint_{S'-\sigma'} \mu \frac{\partial}{\partial \nu} \frac{1}{r} \sec \gamma \, dS' = \lim \iint_{S'-\sigma'} \mu \frac{\zeta - \xi\phi_\xi - \eta\phi_\eta}{r^3} dS'$$

exists, σ' being a circle about p, whose radius tends to 0. But since μ is bounded and integrable,

$$\lim \iint_{S'-\sigma'} \mu \frac{\zeta}{r^3} dS'$$

exists, by Lemma 5. Moreover,

$$\iint_{S'-\sigma'} \left| \mu \frac{- \xi\phi_\xi - \eta\phi_\eta}{r^3} \right| dS' \leqq \max |\mu| \int_0^{2\pi} \int_\epsilon^a \frac{|\phi_\xi| + |\phi_\eta|}{r'} dr' d\theta$$

where ϵ is the radius of σ', and a the maximum radius of S'. As the integral on the right converges, because S is in $C^{1+\delta}$, the rest of the integral under discussion converges absolutely. Thus the integral for U has a meaning at points of S. We denote its value by U_0.

If we introduce $\cos\alpha$, $\cos\beta$, $\cos\gamma$, the direction cosines of the normal to S at $q(\xi, \eta, \zeta)$, U takes the form

$$U = - \iint_S \mu \cos \alpha \frac{\partial}{\partial x} \frac{1}{r} dS - \iint_S \mu \cos \beta \frac{\partial}{\partial y} \frac{1}{r} dS - \iint_S \mu \cos \gamma \frac{\partial}{\partial z} \frac{1}{r} dS.$$

We suppose S is referred to tangent-normal axes at p. The first two terms on the right then represent tangential derivatives of simple distributions on S, while the third is the normal derivative of such a distribution. If μ is continuous at p, $\mu \cos\alpha$ and $\mu \cos\beta$ satisfy a uniform Dini condition there, as is readily verified, and $\mu \cos\gamma$ is continuous. From these facts and from Theorems I and II we obtain

THEOREM IV. *Let S be a regular surface element of class $C^{1+\delta}$. Let μ, the moment of a double distribution on S, be bounded and integrable on S and continuous at the interior point p of S. Then, as P approaches p along the normal to S at p from either side, the potential U of the double distribution approaches limits given by*

$$U_+ = 2\pi\mu(p) + U_0,$$
$$U_- = -2\pi\mu(p) + U_0.$$

On any closed portion of S containing no boundary points of S, on which μ is continuous, these limits are approached uniformly.

By the reasoning used in proving Theorem III we have

THEOREM V. *Let S be a regular surface element of class $C^{1+\delta}$. Let μ, the moment of a double distribution on S, be continuous on S. Then the potential U of the distribution, when defined on S by its limits, is continuous in any region of type V.*

5. **Korn's identities.** In studying the derivatives of order n of the potentials of simple and double distributions we shall use two identities due to Korn.* We shall state these identities in the following lemmas.

LEMMA 6. *If S is a regular surface element of class C'' and σ is continuously differentiable on S, the identity*

$$\begin{aligned}\frac{\partial}{\partial x}\iint_S \sigma \frac{1}{r} dS = &\int_\Gamma \sigma \frac{\cos(\nu, \eta)\cos(s, \zeta) - \cos(\nu, \zeta)\cos(s, \eta)}{r} ds \\ &+ \iint_S \frac{1}{r}\{\sigma_\xi - l[\sigma(l_\xi + m_\eta + n_\zeta) + \sigma_\xi l + \sigma_\eta m + \sigma_\zeta n]\} dS \\ &+ \iint_S \sigma l \frac{\partial}{\partial \nu}\frac{1}{r} dS\end{aligned}$$

(where Γ is the boundary of S, and l, m, n are the direction cosines of the normal at q) *holds for points not on S.*

LEMMA 7. *If S is a regular surface element and μ is continuously differentiable on S, the identity*

$$\begin{aligned}\frac{\partial}{\partial x}\iint_S \mu \frac{\partial}{\partial \nu}\frac{1}{r} dS = &\int_\Gamma \mu \frac{\cos(s, \zeta)\cos(r, \eta) - \cos(s, \eta)\cos(r, \zeta)}{r^2} ds \\ &+ \frac{\partial}{\partial x}\iint_S l\mu_\xi \frac{1}{r} dS + \frac{\partial}{\partial y}\iint_S l\mu_\eta \frac{1}{r} dS + \frac{\partial}{\partial z}\iint_S l\mu_\zeta \frac{1}{r} dS \\ &+ \iint_S \mu_\xi \frac{\partial}{\partial \nu}\frac{1}{r} dS\end{aligned}$$

(where Γ is the boundary of S, and l, m, n are the direction cosines of the normal at q) *holds for points not on S.*

---

* A. Korn, *Potentialtheorie*, vol. I, Berlin, 1899, pp. 36–38, pp. 40–42.

6. **Existence and continuity of the derivatives of order n of the potentials of simple and double distributions.** We have already obtained, in Theorem III, sufficient conditions that the derivatives of the first order of a simple distribution on S be continuous when defined on S by their limiting values. Using Theorems III and V and Lemma 7, we find that if S is a regular surface element of class $C^{1+\delta}$ and μ is of class $C^{1+\delta}$ on S, the potential of the double distribution of moment μ on S is continuously differentiable in a *region of type V*.

We shall now prove the general theorems for derivatives of order n.

THEOREM VI. *Let S be a regular surface element of class $C^{n+\delta}$. Let σ, the density of a simple distribution on S, be of class $C^{n-1+\delta}$. Then U, the potential of this distribution, has continuous derivatives of order n in any region of type V when they are defined on S by their limits.*

THEOREM VII. *Let S be a regular surface element of class $C^{n+\delta}$. Let μ, the moment of a double distribution on S, be of class $C^{n+\delta}$. Then U, the potential of this distribution, has continuous derivatives of order n in any region of type V when they are defined on S by their limits.*

We prove these theorems, already established for $n=1$, by induction. We assume that both theorems are true when n is replaced by $n-1$.

Let S have the standard representation $\zeta=\phi(\xi, \eta)$ with respect to the (ξ, η, ζ)-axes. Then ϕ is of class $C^{n+\delta}$. For the potential of Theorem VI, by Lemma 6, we have

$$
\begin{aligned}
\frac{\partial U}{\partial x} = & \int_\Gamma \sigma \frac{\cos(\nu, \eta)\cos(s, \zeta) - \cos(\nu, \zeta)\cos(s, \eta)}{r} ds \\
& + \iint_S \{\sigma_\xi - l[\sigma(l_\xi + m_\eta) + \sigma_\xi l + \sigma_\eta m]\} \frac{1}{r} dS \qquad (5) \\
& + \iint_S l\sigma \frac{\partial}{\partial \nu} \frac{1}{r} dS,
\end{aligned}
$$

with similar expressions for $\partial U/\partial y$ and $\partial U/\partial z$.

The first term in (5) is analytic at all points not on Γ. Therefore, this term has continuous derivatives of all orders in the closed region V. The second term is the potential of a simple distribution on S with density of class $C^{n-2+\delta}$. By our assumption that the theorem is true when n is replaced by $n-1$, this term has continuous derivatives of order $n-1$ in V when they are defined on S by their limits. The same is true of the third term, since it is the potential of a double distribution on S with moment of class $C^{n-1+\delta}$. Therefore, the whole expression (5) has continuous derivatives of order $n-1$ in V

when they are defined on S by their limits. The expressions for the other two derivatives of U can be discussed in a similar manner, with the result that they also have continuous derivatives of order $n-1$ in V. Therefore, U has continuous derivatives of order n in V when they are defined on S by their limiting values.

Turning to the proof of Theorem VII, we study the potential

$$U = \iint_S \mu \frac{\partial}{\partial \nu} \frac{1}{r} dS$$

for which, by Lemma 7,

$$\begin{aligned} \frac{\partial U}{\partial x} &= \int_\Gamma \mu \frac{\cos(s,\zeta)\cos(r,\eta) - \cos(s,\eta)\cos(r,\zeta)}{r^2} ds \\ &\quad + \frac{\partial}{\partial x}\iint_S l\mu_\xi \frac{1}{r} dS + \frac{\partial}{\partial y}\iint_S l\mu_\eta \frac{1}{r} dS + \iint_S \mu_\xi \frac{\partial}{\partial \nu}\frac{1}{r} dS, \end{aligned} \tag{6}$$

with similar expressions for $\partial U/\partial y$ and $\partial U/\partial z$.

The reasoning employed for Theorem VI now yields the desired result.

We have proved then, that if the theorems are true for derivatives of order $n-1$, they are true for derivatives of order n. We know that they hold for $n=1$. Therefore, they hold for any positive integral n.

7. **On the scope of the sufficient conditions established above.** We have just found sufficient conditions that surface potentials have continuous derivatives of order n. It is natural to ask if we can place lighter conditions upon the spreads without impairing the conclusions of Theorems VI and VII. If we can produce examples satisfying slightly lighter hypotheses and show that the derivatives of order n do not exist or are not continuous, it follows that these lighter hypotheses cannot furnish sufficient conditions for the existence and continuity of derivatives of order n.

Let us consider the derivatives of order n of a simple distribution. Is it possible to get a set of conditions for the existence and continuity of these derivatives in which we require of σ only that it be of class C^{n-1}?

We let S be a circular lamina of radius $a<1$. The plane of S will be the (ξ, η)-plane. We write $\xi = r' \cos\theta$, $\eta = r'\sin\theta$. We consider the densities

$$\sigma = -\frac{r'^{n-1}\cos\theta}{\log r'}, \text{ if } n \text{ is odd}, \quad \sigma = -\frac{r'^{n-1}}{\log r'}, \text{ if } n \text{ is even},$$

where r' is measured from the center of S. We see that σ has continuous derivatives of order $n-1$ on S and that these derivatives do not satisfy a Dini condition at $r'=0$. We consider

$$\left.\frac{\partial^n U}{\partial x^n}\right|_P$$

where P is the point $(0, 0, z)$, $z \neq 0$, that is, a point on the normal to S at the center p. We shall prove that this derivative becomes infinite as P approaches p.

In fact,

$$\left.\frac{\partial^n U}{\partial x^n}\right|_P = \iint_S \sigma \frac{\partial^n}{\partial x^n} \frac{1}{r} dS,$$

and

$$\left.\frac{\partial^n}{\partial x^n} \frac{1}{r}\right|_{x=y=0} = (-1)^n \frac{C_n P_n(u)}{r^{n+1}}, \quad u = -\frac{\xi}{r}, \quad r^2 = r'^2 + z^2,$$

where $C_n > 0$, and where $P_n(u)$ is the Legendre polynomial of order n. Then

$$\left.\frac{\partial^n U}{\partial x^n}\right|_P = (-1)^n C_n \iint_S \sigma \frac{P_n(u)}{r^{n+1}} dS$$

$$= C_n \iint_S \sigma \frac{P_n\left(\frac{\xi}{r}\right)}{r^{n+1}} dS.$$

If we break S into the circle of radius αz and the remaining annular region, where α is a suitably chosen constant and $\alpha z < a$, it can be shown that, when the density σ is the function given above, the integral over the first of these regions approaches 0 with z while the integral over the second region becomes infinite. Hence, the derivative in question becomes infinite as P approaches p.

Therefore, σ being of class C^{n-1} does not insure the existence of the derivatives of order n no matter how smooth the surface S.

By similar reasoning it can be shown that S being of class C^n does not insure the existence of the derivatives of order n of a simple distribution on S no matter how smooth the density. In proving this we consider the distribution of unit density on the surface given by

$$\zeta = \frac{r'^n}{\log r'}, \text{ if } n \text{ is odd}, \quad \zeta = \frac{r'^n \cos\theta}{\log r'}, \text{ if } n \text{ is even},$$

where

$$r'^2 = \xi^2 + \eta^2, \qquad \xi = r' \cos\theta, \qquad \eta = r' \sin\theta.$$

Turning now to double distributions, it can be shown, by considering the distribution on a circular lamina with moment μ defined by

$$\mu = -\frac{r'^n}{\log r'}, \text{ if } n \text{ is odd}, \mu = -\frac{r'^n \cos\theta}{\log r'}, \text{ if } n \text{ is even},$$

that μ being of class C^n does not insure the existence of the derivatives of order n of the potential of a double distribution on S no matter how smooth S.

Using a distribution of moment ξ on the surface given by

$$\zeta = \frac{r'^n}{\log r'}, \text{ if } n \text{ is odd}, \zeta = \frac{r'^n \cos\theta}{\log r'}, \text{ if } n \text{ is even},$$

it can be shown that S being of class C^n does not insure the existence of the derivatives of order n of the potential of a double distribution on S, no matter how smooth the moment.

It thus appears unlikely that materially lighter conditions exist which are at once simple and sufficient for the existence and continuity of the derivatives.

8. **Hölder conditions on the derivatives of surface distributions.** It has been proved* that if S is of class $C^{1+\lambda}$ and σ is of class $C^{\lambda'}$, the derivatives of the first order of the simple distribution of density σ on S satisfy a uniform Hölder condition, in a closed *region of type V*, with exponent κ, where $0<\kappa \leqq \lambda$, $\kappa \leqq \lambda'$. If S is of class $C^{1+\lambda}$ and μ, the moment of a double distribution on S, is of class $C^{1+\lambda'}$, the derivatives of the first order of the potential of this distribution satisfy a uniform Hölder condition in a *region of type V* with exponent κ. We shall now establish general theorems for Hölder conditions on the derivatives of order n.

Theorem VIII. *Let S be a regular surface element of class $C^{n+\lambda}$. Let σ, the density of a simple distribution on S, be of class $C^{n-1+\lambda'}$. Then the derivatives of order n of the potential of this distribution when defined on S by their limiting values satisfy a uniform Hölder condition in any region of type V with exponent κ, where* $0<\kappa \leqq \lambda$, $\kappa \leqq \lambda'$.

Theorem IX. *Let S be a regular surface element of class $C^{n+\lambda}$. Let μ, the moment of a double distribution on S, be of class $C^{n+\lambda'}$. Then the derivatives of order n of the potential of this distribution when defined on S by their limiting values satisfy a uniform Hölder condition in any region of type V with exponent κ, where* $0<\kappa \leqq \lambda$, $\kappa \leqq \lambda'$.

---

* See Schauder, loc. cit. Although in the case of the derivatives of the first order, the present methods do not yield more general results than those of Schauder, in the case of the potential of a double distribution itself, they yield a lighter condition on the surface than that employed by him. See Theorem X, p. 156.

As in an earlier instance, we use the method of induction, assuming that both theorems are true when n is replaced by $n-1$. Then, turning first to the proof of Theorem VIII for $n=n$, we consider the identity (5) for $\partial U/\partial x$.

The first term is analytic at all points not on Γ, and so in V. The second term is the potential of a simple distribution on S with density of class $C^{n-2+\kappa}$. Therefore, this potential has continuous derivatives of order $n-1$ which satisfy a uniform Hölder condition in V with exponent κ. The third term is the potential of a double distribution on S with moment of class $C^{n-1+\kappa}$. Therefore, the derivatives of order $n-1$ of this potential satisfy a uniform Hölder condition in V with exponent κ. Hence, the derivative of U with respect to x has continuous derivatives of order $n-1$ which satisfy a uniform Hölder condition in V with exponent κ. The same is true of the other two derivatives of U. Therefore, the derivatives of U of order n satisfy a uniform Hölder condition in V with exponent κ.

Similar reasoning applied to (6) yields the desired result in the case of Theorem IX.

Since both theorems have been established for $n=1$, it follows that they hold for any positive integral n.

9. Hölder conditions on the potentials of surface distributions. We have been concerned with the existence of Hölder conditions on the *derivatives* of potentials of simple and double distributions on regular surfaces. We now take up the question of such conditions on the potentials themselves.

For a simple distribution, it is known that a distribution with bounded integrable density σ, on a regular surface element S, has a potential which satisfies a uniform Hölder condition with any given exponent less than 1. In fact, the present methods yield the result for such a potential

$$| U_2 - U_1 | \leqq 2\pi \max | \sigma \sec \gamma | r_{12} \log \frac{Re^6}{r_{12}}, \quad r_{12} \leqq R,$$

where R is the maximum chord of S.

We therefore turn at once to the potential of a double distribution, establishing first

LEMMA 8. *Let S be a regular surface element of class $C^{1+\delta}$. Then the integral*

$$\iint_S \left| \frac{\partial}{\partial \nu} \frac{1}{r} \right| dS \tag{7}$$

*is bounded in any region of type V.**

* Schauder (loc. cit.) has proved a similar lemma, less general than this, in that S is required to be of class $C^{1+\lambda}$.

As we have seen in connection with the proof of Theorem III, there is a neighborhood N of any interior closed portion of S with the property that through any point of N there passes a normal to S at an interior point of S. This neighborhood can be so chosen as to include all points of S in V. The points of V not in N are distant at least δ from S, δ being a positive constant. At such points the integral (7) is bounded. In fact

$$\frac{1}{\delta^2}\iint_S dS$$

is such a bound.

We may assume then that the parameter point P in (7) is on the normal to S at the interior point p. We take a tangent-normal system of axes at p and denote by σ a portion of S having a standard representation with these axes. This representation exists uniformly as to p. Then

$$\iint_{S-\sigma}\left|\frac{\partial}{\partial\nu}\frac{1}{r}\right| dS$$

is bounded uniformly as to p.

For the rest of the integral, we have

$$\iint_\sigma \left|\frac{\partial}{\partial\nu}\frac{1}{r}\right| dS = \iint_{\sigma'}\frac{|z-\zeta+\phi_\xi\xi+\phi_\eta\eta|}{r^3}\,dS'$$
$$\leqq I_1 + I_2 + I_3,$$

where

$$I_1 = \iint_{\sigma'}\frac{|z|}{\rho^3}\,dS',\quad I_2 = \iint_{\sigma'}|z|\left|\frac{1}{r^3}-\frac{1}{\rho^3}\right| dS',$$
$$I_3 = \iint_{\sigma'}\frac{|-\zeta+\phi_\xi\xi+\phi_\eta\eta|}{r'^3}\,dS',$$
$$r'^2 = \xi^2+\eta^2,\quad \rho^2 = r'^2+z^2,$$

σ' being the projection of σ on the tangent plane at p. Then, since

$$\iint_{\sigma'}\frac{z}{\rho^3}\,dS' = \Omega$$

where Ω is the solid angle subtended at $P(0, 0, z)$ by the flat element of surface σ',

$$I_1 \leqq 2\pi.$$

Also,

$$I_2 \leqq 3 \iint_{\sigma'} \frac{|\zeta|}{r'^3} dS'.$$

By Lemma 5, this integral is bounded and the bound can be taken independent of p. The integral I_3 is bounded, as we have noted at the beginning of §4. Therefore,

$$\iint_{\sigma} \left| \frac{\partial}{\partial \nu} \frac{1}{r} \right| dS$$

is bounded uniformly as to p. It follows that (7) is bounded in V.

We now prove

THEOREM X. *Let S be a regular surface element of class $C^{1+\delta}$. Let μ, the moment of a double distribution on S, be of class C^{λ}. Then the potential of this distribution, when defined on S by its limits, satisfies a uniform Hölder condition with exponent λ in any region of type V.*

As in Lemma 8, we shall restrict ourselves to points of V in the neighborhood N, since the potential is analytic in the rest of V. We let P_1 be a point of V on the normal to S at the interior point p_1. We shall now assume that all of S is given by the portion contained in a sphere of radius c about p_1 and having a standard representation with tangent-normal axes at p_1, since the potential of the rest of S will be analytic at P_1. Such a representation exists uniformly as to p_1. We let P_2 be a second point of V so restricted that the parallel through P_2 to P_1p_1 meets S in the interior point p_2.

We write

$$U = U_1 + U_2$$

where

$$U_1 = \mu(p_1) \iint_S \frac{\partial}{\partial \nu} \frac{1}{r} dS, \quad U_2 = \iint_S [\mu(q) - \mu(p_1)] \frac{\partial}{\partial \nu} \frac{1}{r} dS.$$

Since U_1 is the potential of a double distribution of constant moment on S, it has continuous derivatives of the first order in V, by Theorem VII.

We have then only to prove that U_2 satisfies a uniform Hölder condition with exponent λ. We denote by σ the portion of S in a circular cylinder of radius $2r_{12}$ whose axis is the normal to S at p_1, r_{12} being the distance $\overline{P_1P_2}$, and write

$$U_2(P_2) - U_2(P_1) = \Delta_1 + \Delta_2,$$

where

$$\Delta_1 = \iint_\sigma [\mu(q) - \mu(p_1)]\left[\frac{\partial}{\partial\nu}\frac{1}{r}\right]_2 dS - \iint_\sigma [\mu(q) - \mu(p_1)]\left[\frac{\partial}{\partial\nu}\frac{1}{r}\right]_1 dS,$$

$$\Delta_2 = \iint_{S-\sigma} [\mu(q) - \mu(p_1)]\left[\frac{\partial}{\partial\nu}\frac{1}{r}\right]_2 dS - \iint_{S-\sigma} [\mu(q) - \mu(p_1)]\left[\frac{\partial}{\partial\nu}\frac{1}{r}\right]_1 dS,$$

the subscripts indicating that the coördinates of P_1 and P_2 are to be substituted for x, y, and z. Then

$$|\Delta_1| \leqq \max|\mu(q) - \mu(p_1)|\left\{\iint_\sigma\left|\left|\frac{\partial}{\partial\nu}\frac{1}{r}\right|_2\right| dS + \iint_\sigma\left|\left|\frac{\partial}{\partial\nu}\frac{1}{r}\right|_1\right| dS\right\}.$$

Since the integral in the above inequality is bounded in V, by Lemma 8, and since μ satisfies a uniform Hölder condition on S, it follows that

$$|\Delta_1| \leqq K_1 r_{12}^{\lambda},$$

K_1 being an appropriate constant.

The remaining difference, Δ_2, may be written in the form

$$\Delta_2 = \int_{s_1}^{s_2}\iint_{S-\sigma}[\mu(q) - \mu(p_1)]\frac{\partial^2}{\partial s\partial\nu}\frac{1}{r}\, dS ds,$$

where the integration with respect to s is from P_1 to P_2 along the segment joining them, and where the integrand is continuous in the field of integration. From this it follows that

$$|\Delta_2| \leqq 10(1 + 2M)A\int_{s_1}^{s_2}\iint_{S-\sigma} r_1^{\lambda}\frac{1}{r^3}\, dS ds,$$

where M is the maximum of $|\phi_\xi|$ and $|\phi_\eta|$ and A is the coefficient in the Hölder condition on μ. Here r is measured from a point on P_1P_2, and r_1 from p_1. If we denote by r_1' and r' the projections on the tangent plane at p_1 of r_1 and r respectively, we have the inequalities

$$r_1 \leqq \max \sec \gamma r_1', \quad r' \leqq r, \quad r_1' \leqq 2r'.$$

Hence,

$$
\begin{aligned}
|\Delta_2| &\leqq 10(1+2M)A\,(\max\sec\gamma)^\lambda \int_{s_1}^{s_2}\iint_{S-\sigma} r_1'^{\lambda-3}\left(\frac{r_1'}{r'}\right)^3 dS\,ds \\
&\leqq 80(1+2M)A\,(\max\sec\gamma)^\lambda \int_{s_1}^{s_2}\iint_{S-\sigma} r_1'^{\lambda-3} dS\,ds \\
&\leqq 80(1+2M)A\,(\max\sec\gamma)^\lambda \int_{s_1}^{s_2}\int_0^{2\pi}\int_{2r_{12}}^{a} r_1'^{\lambda-2} dr'\,d\theta\,ds \\
&\leqq \frac{160(1+2M)\pi}{2^{1-\lambda}(1-\lambda)} A\,(\max\sec\gamma)^\lambda r_{12}^\lambda \\
&\leqq K_2 r_{12}^\lambda.
\end{aligned}
$$

Combining the inequalities for Δ_1 and Δ_2, we see that U_2 satisfies a uniform Hölder condition with exponent λ in V. It follows that the same is true of U.

10. **The existence and continuity of derivatives of volume potentials.** The derivatives of the first order of the potential due to a volume distribution of bounded and integrable density in a bounded volume are continuous throughout space as is well known. As a preliminary to the study of the derivatives of higher order, we define the Dini condition in a region of space.

Definition. Let $f(Q)$ be defined in the closed region V, and continuous at the point P of V. Let Q vary along a ray through P. Then if the integral

$$
\int_0^a \frac{|f(Q)-f(P)|}{r}\,dr,\quad r=\overline{PQ}, \tag{8}
$$

converges at P uniformly as to the direction of the ray chosen, $f(Q)$ is said to satisfy *a uniform Dini condition at P*. If $f(Q)$ is continuous in V and if the integral (8) converges at every point of V uniformly both as to P and as to the direction of the integration at P, $f(Q)$ is said to satisfy *a uniform Dini condition in V*.

In case P is a boundary point of V, a given ray may contain both points of V and points not in V. In this case, the above integral is to be understood as the Lebesgue integral over the closed set of points of the interval $(0, a)$, of the ray in V. Since the integrand is non-negative, the convergence of the integral is equivalent to the summability of the integrand, and the uniformity of the convergence means that the integral vanishes with a, uniformly as to the direction of the ray. The condition at a boundary point is evidently fulfilled if it is possible so to extend the definition of $f(Q)$ to a neighborhood of the boundary point that it is fulfilled by the extended function.

We shall adopt for functions defined in a region of space the class notations given in §2 for a function defined on a surface.

Our object is the establishment of the following theorem.

THEOREM XI. *Let V be a regular closed region of space whose boundary S contains a regular surface element Σ of class $C^{n-1+\delta}$, and let κ, the density of a distribution in V, be of class $C^{n-2+\delta}$, $n\geqq 2$. If $n>2$ let κ be of class $C^{n-2+\delta}$ on Σ as well as in V.* Let V' be a closed region of space partly bounded by Σ, but containing no boundary points of Σ or other points of S. Then the potential of this distribution has continuous derivatives of order n in V' when they are defined on Σ by their limits.*

We note that the region V', except for boundary points in Σ, may be entirely interior to V, or entirely exterior to V. The two cases require different treatment and will therefore be taken up separately. We remark that we may once and for all confine ourselves to the portion of V inside a certain sphere which cuts off from Σ a portion having a standard representation, $\zeta=\phi(\xi,\eta)$, where $\phi(\xi, \eta)$ is in class $C^{n-1+\delta}$, when referred to axes tangent and normal to Σ at the center of the sphere. For later purposes we note that a single radius will serve for this sphere for all points on any closed portion of Σ containing no boundary points of Σ. The justification of confining ourselves to such a portion of V lies first in the fact that the distribution on the rest of V is analytic at the interior points of the sphere, and secondly that exactly the same methods as those here used may be applied to points of V' farther from Σ.

We shall consider then a region V bounded by a spherical surface and a portion of Σ. We prove the theorem first for $n=2$ and then extend it by induction. We may confine ourselves to one derivative of the second order, since the others may be treated by exactly the same methods.

Considering first the case where V' is in V, we take a point P_1 in V', and denote by κ_1 the density at P_1. We break up the integral into two parts,

$$U=\iiint_V \kappa\frac{1}{r}\,dV = U_1+U_2,$$

where

---

* A function defined in a volume may satisfy a uniform Dini condition at a boundary point of the volume, and yet fail to satisfy a Dini condition, as defined for functions of position on a surface, at this point of the bounding surface. For example, the values on the surface of the function

$$f(Q)=\frac{1}{\log\log\dfrac{1}{\zeta}\,\log\dfrac{1}{\zeta}},\quad \zeta\geqq 0,$$

$$=0,\qquad \zeta\leqq 0,$$

considered in the volume bounded by $\zeta=r'^2$, $r'=a$, $a<1$, and $\zeta=-1$, where $r'^2=\xi^2+\eta^2$, fail to satisfy a Dini condition at the origin, although the integral taken along a ray converges uniformly there.

$$U_1 = \kappa_1 \iiint_V \frac{1}{r}\, dV, \quad U_2 = \iiint_V (\kappa - \kappa_1)\frac{1}{r}\, dV.$$

The derivative of the first may be written, with the help of Green's Theorem,

$$\frac{\partial^2 U_1}{\partial x^2} = -\kappa_1 \frac{\partial}{\partial x} \iiint_V \frac{\partial}{\partial \xi} \frac{1}{r}\, dV = \kappa_1 \frac{\partial}{\partial x} \iint_S \cos(\nu, x) \frac{1}{r}\, dS.$$

The differentiation of U_2 at the point P_1 may in the present case be carried out under the integral sign because of the sufficiently rapid vanishing of $(\kappa - \kappa_1)$.* We may therefore write

$$(9) \qquad \left.\frac{\partial^2 U}{\partial x^2}\right|_1 = \kappa_1 \left[\frac{\partial}{\partial x} \iint_S \cos(\nu, x) \frac{1}{r}\, dS\right]_1 + \iiint_V (\kappa - \kappa_1) \left[\frac{\partial^2}{\partial x^2} \frac{1}{r}\right]_1 dV,$$

the subscript indicating that the coördinates of the corresponding point are to be substituted for x, y, and z.

Let P_2 be a second point near P_1. We write

$$\left.\frac{\partial^2 U}{\partial x^2}\right|_2 - \left.\frac{\partial^2 U}{\partial x^2}\right|_1 = \Delta_1 + \Delta_2,$$

where

$$\Delta_1 = \kappa_2 \left[\frac{\partial}{\partial x} \iint_S \cos(\nu, x) \frac{1}{r}\, dS\right]_2 - \kappa_1 \left[\frac{\partial}{\partial x} \iint_S \cos(\nu, x) \frac{1}{r}\, dS\right]_1,$$

$$\Delta_2 = \iiint_V (\kappa - \kappa_2) \left[\frac{\partial^2}{\partial x^2} \frac{1}{r}\right]_2 dV - \iiint_V (\kappa - \kappa_1) \left[\frac{\partial^2}{\partial x^2} \frac{1}{r}\right]_1 dV.$$

The integral in Δ_1 is the potential of a simple spread on S whose density $\cos(\nu, x)$ is in C^4 on Σ, and hence, by Theorem III, it has continuous derivatives in V'. Since κ is also continuous, there corresponds to any $\epsilon > 0$ a $\delta' > 0$ depending on ϵ alone, such that for $\overline{P_1 P_2} < \delta'$,

$$|\Delta_1| < \epsilon/4.$$

In the discussion of Δ_2, we introduce a sphere σ, of radius a about P_1, and impose a second restriction $\overline{P_1 P_2} < a/2$, on P_2. Then, because of the Dini condition on κ, uniform in V, and the fact that

$$\left| \frac{\partial^2}{\partial x^2} \frac{1}{r} \right| \leqq \frac{2}{r^3},$$

it is possible to restrict a so that

* Petrini, loc. cit., p. 135.

$$\left|\iiint_\sigma (\kappa - \kappa_1)\left[\frac{\partial^2}{\partial x^2}\frac{1}{r}\right]_1 dV\right| < \epsilon/4,$$

$$\left|\iiint_\sigma (\kappa - \kappa_2)\left[\frac{\partial^2}{\partial x^2}\frac{1}{r}\right]_2 dV\right| < \epsilon/4,$$

independently of the position of P_2 in the sphere $\overline{P_1P_2}<a/2$. Thus if we write

$$\Delta_2 = \Delta_{21} + \Delta_{22},$$

in the first term of which the field of integration is σ, we have

$$|\Delta_{21}| < \epsilon/2.$$

The second term, Δ_{22}, is the difference in values at P_2 and P_1 of an integral whose integrand is continuous in the variables of integration and the coördinates of the parameter point for $\overline{P_1P_2}<a/2$. Hence, if a is fixed, there exists a $\delta''>0$ such that if $\overline{P_1P_2}<\delta''$,

$$|\Delta_{22}| < \epsilon/4.$$

Thus, if a is a fixed positive number, such that $|\Delta_{21}|<\epsilon/2$, which we have seen is possible, independently of P_2, and if δ is the least of the positive numbers δ', $a/2$, δ'', then for $\overline{P_1P_2}<\delta$,

$$\left|\left.\frac{\partial^2 U}{\partial x^2}\right|_2 - \left.\frac{\partial^2 U}{\partial x^2}\right|_1\right| \leqq |\Delta_1| + |\Delta_{21}| + |\Delta_{22}| < \epsilon.$$

Accordingly, the continuity of this derivative in V' is established in the case in which V' is in V.

Suppose now that V' is exterior to V except for common boundary points which are interior points of Σ. We shall show that the derivative in question approaches limits uniformly along normals to Σ on any closed portion of Σ including no boundary points of Σ. The reasoning of Theorem III will then establish the continuity of this derivative in V'.

Let p be an interior point of Σ, and let P be a point in V' on the normal to Σ at p. If the (x, y)-plane is chosen tangent to Σ at p, with the positive z-axis through P, the value of the derivative at P is given by

$$\frac{\partial^2 U}{\partial x^2} = \iiint_V (\kappa - \kappa_0)\frac{\partial^2}{\partial x^2}\frac{1}{r}\,dV + \kappa_0\frac{\partial}{\partial x}\iint_S \cos(\nu, x)\frac{1}{r}\,dS, \tag{10}$$

where κ_0 is the density at p. Here, the second term is continuous in V' as we have already seen, and so approaches a limit as P approaches p along the normal, uniformly in V'. The first term we write as I_1+I_2 where

$$I_1 = \iiint_\sigma (\kappa - \kappa_0)\frac{\partial^2}{\partial x^2}\frac{1}{r}\,dV, \quad I_2 = \iiint_{V-\sigma} (\kappa - \kappa_0)\frac{\partial^2}{\partial x^2}\frac{1}{r}\,dV,$$

σ being the portion of V in a sphere of radius a about p. A first restriction on a is that it be less than the distance from p to the boundary of Σ. For the first integral, we have

$$\begin{aligned} |I_1| &= \left|\iiint_\sigma (\kappa - \kappa_0)\left[\frac{3(\xi - x)^2}{r^5} - \frac{1}{r^3}\right] dV\right| \leqq 2\iiint_\sigma \frac{|\kappa - \kappa_0|}{r^3}\,dV \\ &= 2\iiint_\sigma \frac{|\kappa - \kappa_0|}{r_0^3}\,\frac{r_0^3}{r^3}\,dV, \end{aligned}$$

where r_0 is measured from p to the point of integration, and r is measured from P. The ratio r_0/r will be greatest when r is the projection of r_0 on a plane parallel to the (x, y)-plane, which may occur for points of σ with positive z-coördinates. However, r_0/r will never exceed $k = \max \sec \gamma$, where γ is the angle between two normals to Σ at points of σ. This number k is independent of p. We have then

$$|I_1| \leqq 2k^3 \iiint_\sigma \frac{|\kappa - \kappa_0|}{r_0^3}\,dV,$$

which is convergent, and vanishes uniformly with a, because κ is in C^δ.

Accordingly, a can be chosen so that the oscillation of I_1 can be made arbitrarily small independently of the position of P on the normal. Then, with a fixed, I_2 is continuous at p. Hence, the derivative approaches a limit along the normal. As the inequalities involved can be made independent of p, the approach is uniform in V'.

Theorem XI is thus established for $n = 2$. In extending the proof, we make use of Green's Theorem to write the derivative of the potential

$$U = \iiint_V \kappa \frac{1}{r}\,dV$$

in the form

$$\frac{\partial U}{\partial x} = \iint_S \kappa \cos(\nu, x)\frac{1}{r}\,dS + \iiint_V \kappa_\xi \frac{1}{r}\,dV, \tag{11}$$

valid for points $P(x, y, z)$ not on S for κ in C'. We assume that Theorem XI has been established for derivatives of order $n-1$. The first term in (11) is the potential of a simple distribution on S with density of class $C^{n-2+\delta}$ on Σ. By Theorem VI, this potential has continuous derivatives of order $n-1$ in

V'. By our assumption that the theorem holds when $n-1$ is substituted for n, the second term has continuous derivatives of order $n-1$ in V'. Therefore $\partial U/\partial x$ has continuous derivatives of order $n-1$ in V'. The same applies to $\partial U/\partial y$ and $\partial U/\partial z$, so that all the derivatives of U with respect to x, y, and z, of order n, exist and are continuous in V'.

Since the theorem is true for derivatives of the second order it follows that it is true for derivatives of any higher order.

11. Hölder conditions on the derivatives of volume potentials. The potential of a distribution of bounded and integrable density in a bounded volume V has continuous derivatives of the first order which satisfy a uniform Hölder condition with exponent λ, where λ is any number such that $0<\lambda<1$.* Considering now the derivatives of higher order, we shall prove

THEOREM XII. *Let V be a regular closed region of space whose boundary S contains a regular surface element Σ of class $C^{n-1+\lambda}$, and let κ, the density of a distribution in V, be of class $C^{n-2+\lambda'}$. Let V' be a closed region of space partly bounded by Σ but containing no boundary points of Σ or other points of S. Then the derivatives of order n of the potential of this distribution, when defined on Σ by their limiting values, satisfy a uniform Hölder condition in V' with exponent λ'', where $0<\lambda''\leqq\lambda$, $\lambda''\leqq\lambda'$.*

The proof is analogous to that of Theorem XI. Our induction begins with the case $n=2$, and to fix ideas, we consider the second derivative with respect to x, first for the case in which V' is in V. By equation (9), this derivative, regarded as a function of P, satisfies

$$\frac{\partial^2 U}{\partial x^2} = \kappa(P)\frac{\partial}{\partial x}\iint_S \cos(\nu, x)\frac{1}{r}\,dS + \iiint_V [\kappa(Q) - \kappa(P)]\frac{\partial^2}{\partial x^2}\,\frac{1}{r}\,dV.$$

The derivative in the first term on the right satisfies a uniform Hölder condition in V' with exponent λ, and by the reasoning of Lemma 1, the first term then satisfies a uniform Hölder condition in V' with exponent λ''.

In establishing the same property for the second term we shall need the following lemma: *Let σ be the portion of V in a sphere of radius α about an interior point of V', α being so restricted that σ contains no points of S other than points of Σ. Then any derivative of the second order of the potential*

$$\iiint_\sigma \frac{1}{r}\,dV$$

* See Korn, *Sur les équations de l'élasticité*, Annales de l'Ecole Normale, (3), vol. 24 (1907), p. 28. Although the bounding surfaces of the volumes considered there have bounded curvatures, the reasoning can be extended so as to apply to any bounded volume.

is bounded at the center of the sphere, and this bound depends only on Σ.

If the sphere is entirely interior to V such a bound is seen at once to exist. We assume then that there are points of Σ in the sphere. Through P, the center of the sphere, there will pass a normal to Σ at one of these points p. If we denote by t the larger of the two regions into which the tangent plane at p divides the sphere, we may write

$$\frac{\partial^2}{\partial x^2}\iiint_\sigma \frac{1}{r}\,dV = \frac{\partial^2}{\partial x^2}\iiint_t \frac{1}{r}\,dV + J,$$

where

$$|J| \leqq \iiint \left|\frac{\partial^2}{\partial x^2}\,\frac{1}{r}\right| dV \leqq 2\iiint \frac{1}{r^3}\,dV,$$

the integral being taken over a set of points containing all points in σ, or t, but not in both (i.e. in $\sigma+t-\sigma\cdot t$). But if we refer to tangent-normal axes at p, and denote by ρ the projection of r on the tangent plane this set is surely included between $\zeta=-A\rho^{1+\lambda}$ and $\zeta=+A\rho^{1+\lambda}$. Therefore,

$$|J| \leqq 4\pi\int_0^\alpha\int_{A-\rho^{1+\lambda}}^{A\rho^{1+\lambda}} \frac{1}{r^3}\,d\zeta\rho d\rho \leqq 8\pi k^3\int_0^\alpha\int_0^{A\rho^{1+\lambda}} \frac{1}{r'^3}\,d\zeta\rho d\rho,$$

where $k=\max\sec\gamma$ on Σ, and where r' is measured from p to the point of integration. It follows that

$$|J| \leqq \frac{8\pi A k^3}{\lambda}\,\alpha^\lambda \leqq \frac{8\pi A k^3 R^\lambda}{\lambda},$$

where R is the maximum chord of V.

Considerations of an elementary nature establish the fact that the integral over t has derivatives of the second order which are bounded at P, and that these bounds can be taken independent of α and the position of the dividing plane.

Returning now to the proof of the theorem for $n=2$ and V' in V, we consider the difference of the values of the second term in the above equality at two interior points, P_1 and P_2. If we denote by σ the portion of V in a sphere of radius $2r_{12}$ about P_1, this difference may be written as $\Delta_1+\Delta_2$, where

$$\Delta_1 = \iiint_\sigma [\kappa(Q)-\kappa(P_2)]\left[\frac{\partial^2}{\partial x^2}\,\frac{1}{r}\right]_2 dV - \iiint_\sigma [\kappa(Q)-\kappa(P_1)]\left[\frac{\partial^2}{\partial x^2}\,\frac{1}{r}\right]_1 dV,$$

$$\Delta_2 = \iiint_{V-\sigma} [\kappa(Q) - \kappa(P_2)]\left[\frac{\partial^2}{\partial x^2}\frac{1}{r}\right]_2 dV$$
$$- \iiint_{V-\sigma} [\kappa(Q) - \kappa(P_1)]\left[\frac{\partial^2}{\partial x^2}\frac{1}{r}\right]_1 dV.$$

Then, if A' is the coefficient in the Hölder condition on κ,

$$|\Delta_1| \leqq 2A'\left\{\iiint_\sigma r_1^{\lambda'-3}dV + \iiint_\sigma r_2^{\lambda'-3}dV\right\}.$$

From this follows the inequality

$$\begin{aligned} |\Delta_1| &\leqq 16\pi A' \int_0^{2r_{12}} r_1^{\lambda'-1}dr \\ &\leqq K_1 r_{12}^{\lambda'}, \end{aligned}$$

for the first integral is not greater than the second.*

Passing to Δ_2, we have

$$\begin{aligned} \Delta_2 &= (\kappa_1 - \kappa_2)\left[\frac{\partial^2}{\partial x^2}\iiint_V \frac{1}{r}dV\right]_1 - (\kappa_1 - \kappa_2)\left[\frac{\partial^2}{\partial x^2}\iiint_\sigma \frac{1}{r}dV\right]_1 \\ &\quad + \iiint_{V-\sigma}[\kappa - \kappa_2]\left[\frac{\partial^2}{\partial x^2}\frac{1}{r}\Big|_2 - \frac{\partial^2}{\partial x^2}\frac{1}{r}\Big|_1\right]dV \\ &= \Delta_{21} - \Delta_{22} + \Delta_{23}, \end{aligned}$$

where κ_1 and κ_2 are the values of κ at P_1 and P_2. The coefficients of $(\kappa_1 - \kappa_2)$ in Δ_{21} and Δ_{22} are bounded independently of P_1 and r_{12}, the first by Theorem XI and the second by the lemma just proved. Therefore,

$$|\Delta_{21}| \leqq K_2 r_{12}^{\lambda'},$$
$$|\Delta_{22}| \leqq K_3 r_{12}^{\lambda'}.$$

There remains Δ_{23} which may be written in the form

$$\Delta_{23} = \int_{s_1}^{s_2}\iiint_{V-\sigma}[\kappa - \kappa_2]\frac{\partial^3}{\partial s \partial x^2}\frac{1}{r}\,dV ds,$$

where the integration with respect to s is from P_1 to P_2 along the segment joining them, and where the integrand is continuous in the field of integration. It follows that

---

* See Kellogg, *Foundations of Potential Theory*, loc. cit., p. 148, Lemma III.

$$|\Delta_{23}| \leqq 18A' \int_{s_1}^{s_2} \iiint_{V-\sigma} r_2^{\lambda'} \frac{1}{r^4}\, dV ds,$$

where r is measured from a point of P_1P_2. Using the inequalities $r_2 < 2r_1$, $r \geqq r_1/2$, we have

$$\begin{aligned} |\Delta_{23}| &\leqq 576A' \int_{s_1}^{s_2} \iiint_{V-\sigma} r_1^{\lambda'-4} dV ds \\ &\leqq 2304\pi A' \int_{s_1}^{s_2} \int_{2r_{12}}^{\infty} r_1^{\lambda'-2} dr ds \\ &\leqq \frac{2304\pi A'}{1-\lambda'} r_{12}^{\lambda'-1} \cdot r_{12} \leqq K_4 r_{12}^{\lambda'}. \end{aligned}$$

When the inequalities are assembled it appears that the second term satisfies a uniform Hölder condition in V' with exponent λ'. Therefore, $\partial^2 U/\partial x^2$ satisfies a uniform Hölder condition with exponent λ''. Similar reasoning holds for the other derivatives of second order, and the theorem holds for $n=2$ when V' is in V.

We turn now to the case in which V' is exterior to V save for common boundary points on Σ. As we have seen, there is a neighborhood N of any interior closed portion of Σ with the property that through any point of N there passes a normal to Σ at an interior point of Σ. This neighborhood can be so chosen as to include all points of Σ in V'. We shall confine ourselves to the points common to N and V', for since U is harmonic at all other points of V', its derivatives satisfy Hölder conditions uniformly in this remaining portion of V'.

Let P_1 be a point in V',—because of what has been said we shall from now on omit the specification that it be in N,—on the normal to Σ at p_1. Let P_2 be a second point of V' such that the parallel through P_2 to P_1p_1 meets Σ in the single point p_2, a situation always attainable by a uniform restriction on $\overline{P_1P_2}$. We refer V to tangent-normal axes at p_1 and write

$$\left.\frac{\partial^2 U}{\partial x^2}\right|_2 - \left.\frac{\partial^2 U}{\partial x^2}\right|_1 = \Delta_1 + \Delta_2,$$

where

$$\Delta_1 = \kappa_2 \left[\frac{\partial}{\partial x} \iint_S \cos(\nu, x) \frac{1}{r}\, dS\right]_2 - \kappa_1 \left[\frac{\partial}{\partial x} \iint_S \cos(\nu, x) \frac{1}{r}\, dS\right]_1,$$

$$\Delta_2 = \iiint_V (\kappa - \kappa_2) \left[\frac{\partial^2}{\partial x^2} \frac{1}{r}\right]_2 dV - \iiint_V (\kappa - \kappa_1) \left[\frac{\partial^2}{\partial x^2} \frac{1}{r}\right]_1 dV,$$

κ_1 and κ_2 being the density at p_1 and p_2 respectively.

Since κ is in $C^{\lambda'}$, and since $\cos(\nu, x)$ is in C^{λ} on Σ, we have, by Theorem VIII and by the reasoning of Lemma 1,

$$|\Delta_1| \leqq K_1 r_{12}^{\lambda''},$$

K_1 being an appropriate constant. Here r_{12} is the distance $\overline{P_1P_2}$; and since the integrals are computed at P_1 and P_2 whereas κ_1 and κ_2 are the values of κ at p_1 and p_2, it is important to notice in the establishment of the above inequality that $\overline{p_1p_2}$ does not exceed the projection of $\overline{P_1P_2}$ times $k(=\max\sec\gamma)$, and hence

$$\overline{p_1p_2} \leqq k\overline{P_1P_2}.$$

In considering Δ_2, we note that the points P_1 and P_2 may be thought of as not on Σ, for if a function, continuous in a closed region, satisfies a uniform Hölder condition in the interior, it satisfies the same uniform Hölder condition in the closed region.

Using now a sphere of radius $2r_{12}$ about p_1 and calling the portion of V in this sphere σ, we may write

$$\Delta_2 = \Delta_{21} + \Delta_{22},$$

where in Δ_{21} the field of integration is σ, and in Δ_{22}, $V-\sigma$. Then

$$\begin{aligned}|\Delta_{21}| \leqq 2\iiint_\sigma \frac{|\kappa-\kappa_2|}{\overline{P_2Q}^3}\,dV \\ + 2\iiint_\sigma \frac{|\kappa-\kappa_1|}{\overline{P_1Q}^3}\,dV,\end{aligned}$$

Q being the point of integration. If we let $r_1 = p_1Q$, and $r_2 = p_2Q$, we have

$$|\Delta_{21}| \leqq 2A'k^3\left\{\iiint_\sigma r_2^{\lambda'-3}dV + \iiint_\sigma r_1^{\lambda'-3}dV\right\},$$

where A' is the coefficient in the Hölder condition on κ. From this follows the inequality

$$\begin{aligned}|\Delta_{21}| &\leqq 16\pi A'k^3\int_0^{2r_{12}} r_1^{\lambda'-1}dr \\ &\leqq K_2 r_{12}^{\lambda'},\end{aligned}$$

since the first integral is not greater than the second.

Passing to Δ_{22}, we have

$$\Delta_{22} = \iiint_{V-\sigma} (\kappa - \kappa_2)\left[\frac{\partial^2}{\partial x^2}\frac{1}{r}\right]_2 dV - \iiint_{V-\sigma} (\kappa - \kappa_1)\left[\frac{\partial^2}{\partial x^2}\frac{1}{r}\right]_1 dV$$

$$= (\kappa_1 - \kappa_2)\iiint_{V-\sigma}\left[\frac{\partial^2}{\partial x^2}\frac{1}{r}\right]_1 dV$$

$$+ \iiint_{V-\sigma} (\kappa - \kappa_2)\left[\frac{\partial^2}{\partial x^2}\frac{1}{r}\bigg|_2 - \frac{\partial^2}{\partial x^2}\frac{1}{r}\bigg|_1\right] dV$$

$$= \Delta_{221} + \Delta_{222}.$$

We may write

$$\Delta_{221} = (\kappa_1 - \kappa_2)\frac{\partial}{\partial x}\iint_{S'} \cos(\nu, x)\frac{1}{r}\,dS,$$

where S' is the boundary of $V-\sigma$, and r is measured from P_1. Since κ satisfies a uniform Hölder condition with exponent λ', Δ_{221} will satisfy an inequality of the desired form if the derivative of the simple spread is bounded. That this is so may be seen by writing it in the form

$$\begin{aligned}\frac{\partial}{\partial x}\iint_{S'} \cos(\nu, x)\frac{1}{r}\,dS &= \frac{\partial}{\partial x}\iint_{S}\cos(\nu, x)\frac{1}{r}\,dS - \frac{\partial}{\partial x}\iint_{\Sigma_1}\cos(\nu, x)\frac{1}{r}dS \\ &\quad + \frac{\partial}{\partial x}\iint_{\Sigma_2} \cos(\nu, x)\frac{1}{r}\,dS,\end{aligned} \tag{12}$$

where Σ_1 is the portion of Σ contained in σ, and Σ_2 is the rest of the boundary of σ.

The first term on the right, being a derivative in a fixed direction of the potential of a simple spread with density of class C^λ, is continuous, and therefore bounded in V'. The same reasoning cannot be applied to the remaining two terms because of the presence of edges at distances from p_1 which approach 0 with r_{12}. But they are, nevertheless, bounded, uniformly as to P_1 and p_1. This may be seen as follows.

In the second term $|\cos(\nu, x)| \leqq A r_1{}^\lambda$, where r_1 is the distance from p_1 to the integration point, because Σ, and hence Σ_1, are of class C^λ, and since $\cos(\nu, x)$ vanishes at p_1 on account of the position of the axes. Hence

$$\left|\frac{\partial}{\partial x}\iint_{\Sigma_1} \cos(\nu, x)\frac{1}{r}\,dS\right| \leqq \iint_{\Sigma_1'} |\cos(\nu, x)|\,\frac{1}{r^2}\sec\gamma\,dS'$$

$$\leqq k^{1+\lambda}A\iint_{\Sigma'} r'^{\lambda-2}dS',$$

where r' is the projection of r and r_1 on the (x, y)-plane, where $k = \max \sec \gamma$ on Σ_1 and the integration is over the projection of Σ_1. The last integral is uniformly convergent and vanishes with r_{12}.

In the third term, we have $2r_{12} \leqq kr$, where $2r_{12}$ is the radius of Σ_2 and $k = \max \sec \gamma$. Hence, the term is not greater in absolute value than

$$\iint_{\Sigma_2} \frac{1}{r^2}\, dS \leqq k^2 \iint_{\Sigma_2} \frac{1}{(2r_{12})^2}\, dS \leqq 4\pi k^2.$$

Thus

$$|\Delta_{221}| \leqq K_3 r_{12}^{\lambda'}.$$

Finally, we have to consider Δ_{222}, which may be written in the form

$$\Delta_{222} = \int_{s_1}^{s_2} \iiint_{V-\sigma} (\kappa - \kappa_2) \frac{\partial^3}{\partial s \partial x^2} \frac{1}{r}\, dV ds,$$

where the integration with respect to s is from P_1 to P_2 along the segment joining them, and where the integrand is continuous in the field of integration.

Let a denote a constant such that in the sphere of radius a about p_1 the angle between any two normals never exceeds $\pi/6$; a can be selected independently of p_1, because of the uniform continuity of the direction cosines of the normal to Σ. If we restrict r_{12} to be less than $a/4$, we may confine ourselves to the portion v of V between the sphere of radius a about p_1, and the sphere σ, of radius $2r_{12}$; for, in the remaining field, the integrand in Δ_{222} is uniformly bounded, and the corresponding integral, accordingly, does not exceed in absolute value a constant times r_{12}. For the rest,

$$\begin{aligned}\left| \int_{s_1}^{s_2} \iiint_{v} (\kappa - \kappa_2) \frac{\partial^3}{\partial s \partial x^2} \frac{1}{r}\, dV ds \right| &\leqq 18A' \int_{s_1}^{s_2} \iiint_{v} r_2^{\lambda'} \frac{1}{r^4}\, dV ds \\ &\leqq 18A' \int_{s_1}^{s_2} \iiint_{v} r_2^{\lambda'-4} \left(\frac{r_2}{r}\right)^4 dV ds,\end{aligned}$$

where r is measured from a point of the segment P_1P_2, and r_2 is measured from p_2. For any fixed position of the point of integration, r_2/r is greatest when r is measured as nearly as possible perpendicular to the normal to Σ at p_1, in which case

$$r_2 \leqq (r + P_1P_2) \max \sec \gamma \leqq \frac{2}{3^{1/2}}(r + r_{12}).$$

Moreover,

$$r \geqq \frac{2r_{12}}{\max \sec \gamma} - r_{12} \geqq (3^{1/2} - 1)r_{12}.$$

Hence,

$$\frac{r_2}{r} \leqq \frac{2}{3^{1/2}}\left(1 + \frac{r_{12}}{r}\right) < 3.$$

Accordingly,

$$\left| \int_{s_1}^{s_2} \iiint_v (\kappa - \kappa_2) \frac{\partial^3}{\partial s \partial x^2} \frac{1}{r} \, dV ds \right| \leqq 1458A' \int_{s_1}^{s_2} \iiint_v r_2^{\lambda' - 4} dV ds$$

$$\leqq 5832\pi A' \int_{s_1}^{s_2} \int_{2r_{12}}^{2a} r_2^{\lambda' - 2} dr_2 ds$$

$$\leqq \frac{5832\pi A'}{1 - \lambda'} r_{12}^{\lambda' - 1} \cdot r_{12}$$

$$\leqq K_4 r_{12}^{\lambda'}.$$

When the inequalities are assembled, it appears that $\partial^2 U/\partial x^2$ satisfies a uniform Hölder condition in V' with exponent λ''. The reasoning requires no modification in the case of other derivatives of the second order, provided one direction is tangential. In the case of $\partial^2 U/\partial z^2$, with the above orientation of the axes, the only point at which modification is necessary is in the proof that the second term in the expression (12) is bounded. For, $\cos(\nu, z) = 1$ at p_1. In this case, however, the term becomes, when we use the projection Σ_1' of Σ_1 on the (x, y)-plane, as the field of integration,

$$\frac{\partial}{\partial z} \iint_{\Sigma_1'} \frac{1}{r} dS' = \iint_{\Sigma_1'} \frac{\zeta - z}{r^3} dS'.$$

This can be shown to be bounded, independently of P_1, p_1 and r_{12}, by comparing it with the same derivative of a spread of unit density on the flat region Σ_1',

$$\iint_{\Sigma_1'} -\frac{z}{\rho^3} dS', \quad \rho^2 = \xi^2 + \eta^2 + z^2,$$

and noting that ρ/r lies uniformly between two positive bounds.

When the details are supplied, the proof of Theorem XII for $n = 2$ is complete.

For $n > 2$, we use the identity (11) and assume that the theorem has been established for derivatives of order lower than n. Then the first term in (11)

has continuous derivatives of order $n-1$ in V' which satisfy there a uniform Hölder condition with exponent λ'', by Theorem VIII. The same is true of the second term, by our assumption about the derivatives of order lower than n. Therefore, the theorem is true for derivatives of order n. Since it has been established for $n=2$, it follows that it holds for any $n \geqq 2$.

12. **Logarithmic potentials.** The potential of a logarithmic distribution on a plane curve can be interpreted as the potential of a distribution on an infinite cylinder with elements perpendicular to the plane of the curve. Furthermore, the potential of the distribution on the portion S of the cylinder, outside two planes parallel to and on either side of the plane of the curve, is continuous and has continuous derivatives of all orders at all points of the plane of the curve.* The situation is the same for logarithmic double distributions and for logarithmic distributions on plane areas.

Knowing this, we can see immediately that all the theorems established in this paper for surface or volume distributions hold also for logarithmic distributions on plane curves or areas, without alteration other than the appropriate changes in dimensionality.

* Kellogg, *Foundations of Potential Theory*, loc. cit., p. 174.

Radcliffe College,
Cambridge, Mass.

THE DEGREE OF CONVERGENCE OF A SERIES OF BESSEL FUNCTIONS*

BY

M. G. SCHERBERG

A number of problems of mathematical physics require the expansion of an arbitrary function in terms of Bessel functions in a manner analogous to the expansion of such a function in trigonometric functions. An important series in Bessel functions, necessary to the solution of one group of problems, the most familiar of which are the problems of the vibrating circular drumhead and the flow of heat in a cylinder, is the Bessel series, which has the form†

$$f(x) = \sum_1^\infty B_n J_0(\lambda_n x)$$

in which the B_n are constants and the λ_n's are the positive roots of the equation

$$l\lambda J_0'(\lambda) + hJ_0(\lambda) = 0$$

where either $l=0$ or $h/l>0$.

The more general series to be studied in this paper has the form

$$f(x) = B(x) + \sum_1^\infty B_n J_\nu(\lambda_n x), \qquad \nu \geqq 0, \tag{1}$$

in which the λ's are the positive roots of the equation‡

$$l\lambda J_\nu'(\lambda) + hJ_\nu(\lambda) = 0 \tag{2}$$

while $B(x)$ is an additional term which is present when (2) has also a pair of imaginary roots $\pm i\lambda_0$. Since the functions $x^{1/2}J_\nu(\lambda_n x)$ form an orthogonal set§ over a range of integration from zero to one, the coefficients B_n are found in the usual formal manner and are

$$B_n = \frac{\int_0^1 xf(x)J_\nu(\lambda_n x)dx}{\int_0^1 xJ_\nu^2(\lambda_n x)dx}. \tag{3}$$

The function $B(x)\equiv 0$ when $l=0$ or $h/l+\nu>0$. Otherwise it has a form depending on whether $h/l+\nu$ is equal to or less than zero.‖

---

* Presented to the Society, September 11, 1931; received by the editors April 25, 1932, and, in revised form, June 20, 1932.

† Watson, *Theory of Bessel Functions*, 1922, pp. 596–597; Byerly, *Fourier's Series*, pp. 12–14.

‡ The notation $J_\nu'(x)$ for $(d/dx)J_\nu(x)$ will be used through the paper.

§ Gray and Mathews, *Treatise on Bessel Functions*, 1922, p. 91.

‖ Watson, loc. cit., pp. 596–597.

That the series converges to the value of the function $f(x)$ under suitable restrictions on the function, and the range of the variable, has been shown by several writers.* The degree of convergence of a series is the order of magnitude of the difference between the function and the first n terms of the series. Thus, if those restrictions are placed upon the function $f(x)$ which insure the convergence of the series to the proper value in a defined range of x, the degree of convergence of the series may be calculated as the order of magnitude of the remainder after n terms.

To avoid undue repetition, a convention of symbol is made at this time. K will designate constants independent of x and n and depending only on such fixed quantities as ν, the number of discontinuities in $f'(x)$, etc. The function θ will be any function of any number of variables which is numerically less than one for all values of the variables considered. The notation $\theta_1(x)$ will indicate a function which has one for an upper bound and which has a bounded derivative with respect to x.

I. The degree of convergence in the absence of higher derivatives

Lemma 1. *If $F(x)/x$ has bounded variation in the interval* $0 \leqq x \leqq 1$, *then*

$$\int_0^1 F(x)J_\nu(\lambda_n x)dx = \frac{K\theta(\lambda_n)}{\lambda_n^{3/2}}. \tag{4}$$

By means of the asymptotic formula†

$$J_\nu(x) = \left(\frac{2}{\pi x}\right)^{1/2}\left\{\cos(x-\alpha) + \frac{K\theta(\nu, x)}{x}\right\}, \tag{5}$$

$$\alpha = \frac{2\nu+1}{4}\pi, \quad \nu \geqq 0,$$

on setting $\Phi = F(x)/x$, we have

$$\int_0^1 F(x)J_\nu(\lambda_n x)dx = \left(\frac{2}{\pi\lambda_n}\right)^{1/2}\int_0^1 \Phi(x)x^{1/2}\cos(\lambda_n x - \alpha)dx + \frac{K\theta}{\lambda_n^{3/2}}.$$

Since $\Phi = \Phi_1(x) - \Phi_2(x)$ in which Φ_1, Φ_2 are monotone increasing, we may assume without loss in generality that Φ is also monotone increasing and hence, by the second law of the mean,

* C. N. Moore, these Transactions, vol. 12 (1911), pp. 181–200; also Watson, loc. cit., pp. 576–605.

† Lipschitz, Crelle's Journal, vol. 56 (1859), pp. 193–196; Watson, loc. cit.; C. N. Moore, loc. cit., p. 189.

$$\int_0^1 \Phi(x)x^{1/2}\cos(\lambda_n x-\alpha)dx$$

$$= \Phi(+0)\int_0^\xi x^{1/2}\cos(\lambda_n x-\alpha)dx + \Phi(1-0)\int_\xi^1 x^{1/2}\cos(\lambda_n x-\alpha)dx$$

$$= \frac{K\theta(\lambda_n)}{\lambda_n}\cdot$$

LEMMA 2. *In the interval* $0\leqq x\leqq 1$ *let the function* $f(x)$ *be absolutely continuous and let* $f'(x)$ *have bounded variation. Then the general coefficient* B_n *of the series* (1) *may be written as*

$$\frac{(2\pi)^{1/2}\delta_l f(1)(-1)^n}{\lambda_n^{1/2}} + \frac{\pi\nu f(0)}{\lambda_n} + \frac{K\theta(\lambda_n)}{\lambda_n^{3/2}}$$

where

$$\delta_l = \begin{cases} 0 & \text{when } l \neq 0, \\ \pm 1 & \text{when } l = 0. \end{cases}$$

The treatment in this part of the paper is similar to that employed by C. N. Moore* in a paper on the uniform convergence of a Bessel series.

The denominator of B_n may be written†

$$\int_0^1 xJ_\nu^2(\lambda_n x)dx = \tfrac{1}{2}\{J_\nu^2(\lambda_n) + J_{\nu+1}^2(\lambda_n)\} - \frac{\nu}{\lambda_n}J_\nu(\lambda_n)J_{\nu+1}(\lambda_n)$$

and with the aid of (5) is readily reduced to the form

$$\int_0^1 xJ_\nu^2(\lambda_n x)dx = \frac{1}{\pi\lambda_n} + \frac{K\theta_1(\lambda_n)}{\lambda_n^2}\cdot \tag{6}$$

By means of (6), B_n now assumes the form

$$[\lambda_n\pi + K\theta_1(\lambda_n)]\int_0^1 xf(x)J_\nu(\lambda_n x)dx. \tag{7}$$

On integration by parts with the aid of the recurrence formula

$$\frac{d}{dx}[x^{\nu+1}J_{\nu+1}(x)] = x^{\nu+1}J_\nu(x), \tag{8}$$

the integral in (7) becomes

* C. N. Moore, loc. cit., p. 183.

† Byerly, *An Elementary Treatise on Fourier Series*, 1902, p. 224, formula 12.

$$\int_0^1 xf(x)J_\nu(\lambda_n x)dx$$

$$= \frac{1}{\lambda_n}\int_0^1 \frac{xf(x)}{(\lambda_n x)^{\nu+1}} d\left[(\lambda_n x)^{\nu+1}J_{\nu+1}(\lambda_n x)\right]$$

$$(9)\qquad = \frac{f(1)J_{\nu+1}(\lambda_n)}{\lambda_n} - \frac{1}{\lambda_n}\int_0^1 xf'(x)J_{\nu+1}(\lambda_n x)dx + \frac{1}{\lambda_n}\int_0^1 \nu f(x)J_{\nu+1}(\lambda_n x)dx$$

$$= \frac{f(1)J_{\nu+1}(\lambda_n)}{\lambda_n} + \frac{\nu f(0)}{\lambda_n}\int_0^1 J_{\nu+1}(\lambda_n x)dx - \frac{1}{\lambda_n}\int_0^1 xf'(x)J_{\nu+1}(\lambda_n x)dx$$

$$+ \frac{\nu}{\lambda_n}\int_0^1 [f(x) - f(0)]J_{\nu+1}(\lambda_n x)dx.$$

If we assume, as we may without loss in generality, that $f'(x)$ is $\geqq 0$ and monotone increasing, then $(1/x)[f(x)-f(0)]$ will be also positive and monotone increasing.

It follows that the last two integrals in the last line of (9) have the form $K\theta(\lambda_n)/\lambda_n^{5/2}$.

The integral $\int_0^1 J_{\nu+1}(\lambda_n x)dx$ of (9) may be written

$$\frac{1}{\lambda_n}\int_0^{\lambda_n} J_{\nu+1}(x)dx = \frac{1}{\lambda_n}\left\{\int_0^\infty J_{\nu+1}(x) - \int_{\lambda_n}^\infty J_{\nu+1}(x)dx\right\}$$

since the integrals in question all converge. Although it is not necessary to go beyond the fact that the integral $\int_0^\infty J_{\nu+1}(x)dx$ is a function of ν alone, its value one* will be utilized. By means of (5)

$$\int_{\lambda_n}^\infty J_{\nu+1}(x)dx = \int_{\lambda_n}^\infty \left(\frac{2}{\pi x}\right)^{1/2} \sin(x-\alpha)dx + K\int_{\lambda_n}^\infty \frac{\theta_1(x)dx}{x^{3/2}} = \frac{K\theta(\lambda_n)}{\lambda_n^{1/2}}.$$

Thus

$$(10)\qquad \int_0^1 J_{\nu+1}(\lambda_n x)dx = \frac{1}{\lambda_n} + \frac{K\theta(\lambda_n)}{\lambda_n^{3/2}}.$$

To complete the proof of Lemma 2, it remains to reduce the term $f(1)J_{\nu+1}(\lambda_n)/\lambda_n$ of (9). A formula for the roots of equation (2) due to Moore† gives

$$(11)\qquad \lambda_n = n\pi + q + \frac{K\theta}{n} = K\gamma(n)\cdot n, \quad 1 \leqq \gamma(n) \leqq 2,$$

* Gray and Mathews, loc. cit., p. 65, Formula 8.

† C. N. Moore, loc. cit., pp. 189–196.

where

$$q = \begin{cases} k\pi - \pi/2 + \dfrac{2\nu + 1}{4}\pi, & l = 0, \\ k\pi + \dfrac{2\nu + 1}{4}\pi, & l \neq 0, \end{cases}$$

and k is an integer, positive, negative or zero.

From (11)

$$\sin(\lambda_n - \alpha) = \sin(n\pi + q - \alpha) + \frac{K\theta(n)}{n}, \qquad \alpha = \frac{2\nu + 1}{4}\pi,$$

$$= \begin{cases} \pm(-1)^n + \dfrac{K\theta(n)}{n}, & l = 0, \\ \dfrac{K\theta(n)}{n}, & l \neq 0, \end{cases}$$

and (5)

$$J_{\nu+1}(\lambda_n) = \delta_l(-1)^n\left(\frac{2}{\pi\lambda_n}\right)^{1/2} + \left(\frac{2}{\pi\lambda_n}\right)^{1/2}\frac{K\theta}{n} + \frac{K\theta}{\lambda_n^{3/2}} \tag{12}$$

in which δ_l is defined as in the statement of Lemma 2. The conclusion of the lemma follows from a combination of the above results.

THEOREM I. *Let $f(x)$ be a function such as described in Lemma 2, and, in addition, let the conditions $\delta_l f(1) = \nu f(0) = 0$ be satisfied. Then*

$$f(x) - S_n(x) = \frac{K\theta(n, x)}{n^{1/2}}, \qquad 0 \leqq x \leqq 1,$$

where $S_n(x) - B(x)$ is the sum of the first n regular terms of the series (1).

It has been shown* that under the conditions of the Lemma 1, the series will converge to the value of the function in any sub-interval of $0 \leqq x \leqq 1$ having zero as an end point provided $f(x)$ is continuous in this sub-interval and the product $\nu f(0) = 0$, and that it will converge to the value of the function in a sub-interval of $0 \leqq x \leqq 1$ having one as an end point if again $f(x)$ is con-

* C. N. Moore, loc. cit., has shown the convergence to $f(x)$ under conditions which insure "closure" and hence the convergence to $f(x)$ under conditions of the lemma follows if there is convergence at all.

tinuous in this sub-interval and the product $\delta_1 f(1)=0$. It, therefore, follows that under the conditions of Theorem I the series will converge to $f(x)$ throughout the interval $0\leqq x\leqq 1$. Further, from Lemma 2 the general term of the series assumes the form

$$\frac{K\theta(\lambda_n)}{\lambda_n^{3/2}}J_\nu(\lambda_n x).$$

Since $J_\nu(\lambda_n x)$ is uniformly bounded,* this general term may be written $K\theta(n)/n^{3/2}$ and the remainder after n terms becomes $K\theta(n)/n^{1/2}$.

THEOREM II. *Let $f(x)$ be a function such as described in Lemma* 1. *Then in the interval* $0\leqq a\leqq x\leqq b\leqq 1$

$$f(x)-S_n(x)=\frac{K\theta(x,n)}{nx^{1/2}}+\frac{K\delta_1 f(1)\theta(n,x)}{(1-x)n^{1/2}}+\frac{K\theta(n,x)}{x^{3/2}n^2}+\frac{\nu f(0)K\theta(x,n)}{x^{3/2}n^{3/2}}$$

where $a\neq 0$ *unless* $\nu f(0)=0$, $b\neq 1$ *unless* $\delta_1 f(1)=0$, *and*

$$\frac{K\theta(n,x)}{nx^{1/2}}=\frac{K\theta(n,x)}{x^{3/2}n^2}\equiv 0 \text{ when } a=0.$$

Under the conditions of Lemma 2, the series (1) becomes†

$$(13)\qquad \begin{aligned}\delta_1 f(1)\sum_{m=n+1}^{\infty}\frac{(-1)^m J_\nu(\lambda_m x)}{\lambda_m^{1/2}}&+\pi\nu f(0)\sum_{m=n+1}^{\infty}\frac{(-1)^m J_\nu(\lambda_n x)}{\lambda_m}\\ &+\sum_{m=n+1}^{\infty}\frac{K\theta(\lambda_m)}{\lambda_m^{3/2}}J_\nu(\lambda_m x).\end{aligned}$$

By means of (5) the general term of the last sum becomes

$$\frac{K\theta(\lambda_m)\cos(\lambda_m x-\alpha)}{\lambda_m^2 x^{1/2}}+\frac{K\theta(\lambda_m)}{x^{3/2}\lambda_m^{3/2}}$$

which with the aid of (11) yields

$$\sum_{m=n+1}^{\infty}\frac{K\theta(\lambda_m)}{\lambda_m^{3/2}}J_\nu(\lambda_m x)=\frac{K\theta(n,x)}{nx^{1/2}}+\frac{K\theta(n,x)}{n^2x^{3/2}},\qquad 0\leqq x\leqq 1.$$

The first sum of (13) is zero when $\nu>0$ and $x=0$ since $J_\nu(0)=0$. On the other hand if $\nu=0$, then $J_\nu(0)=1$, and this sum is that of an alternating

* Watson, loc. cit., p. 44.

† The convergence of the separated parts will be apparent in what follows.

series of numerically decreasing terms. It has a value, then, which is numerically less than the first term or $K\theta(n)/(n+1)^{1/2}$. When $x>0$ the summation is made in three parts.

Let r and s be the smallest integers greater than $(n-1)$ which satisfy the inequalities $\lambda_{r+1}x>k_1$ and $\lambda_{s+1}x>k_2$ where k_1 and k_2 are the first and second positive roots of $J_\nu'(x)=0$. Now when x is small r and s will surely be larger than $(n+1)$ and this sum may be divided into three parts in the first two of which $J_\nu(\lambda_n x)$ is monotone. With the arguments of the $\sum$'s omitted, they are

$$\sum_{n+1}^{r} + \sum_{r+1}^{s} + \sum_{s+1}^{\infty} = \sigma_1 + \sigma_2 + \sigma_3.$$

By means of (5)

$$\begin{aligned}(14)\qquad \sigma_3 &= \left(\frac{2}{\pi x}\right)^{1/2}\left[\sum_{m=s+1}^{\infty}\frac{(-1)^m \cos(\lambda_m x-\alpha)}{\lambda_m} + \frac{1}{x}\sum_{m=s+1}^{\infty}\frac{(-1)^m\theta_1(\lambda_m x)}{\lambda_m^2}\right]\\ &= \sigma_4+\sigma_5.\end{aligned}$$

The sum σ_5 converges absolutely, hence

$$\left(\frac{\pi}{2}\right)^{1/2}\sigma_5 = \pm\, x^{-3/2}\sum_{k=s+1,s+3,\cdots}\left\{\frac{\theta_1(\lambda_k x)}{\lambda_k^2} - \frac{\theta_1(\lambda_{k+1}x)}{\lambda_{k+1}^2}\right\}.$$

But

$$\frac{\theta_1(\lambda_k x)}{\lambda_k^2} - \frac{\theta_1(\lambda_{k+1}x)}{\lambda_{k+1}^2} = \frac{\theta_1(\lambda_k x)-\theta_1(\lambda_{k+1}x)}{\lambda_k^2} + \theta_1(\lambda_{k+1}x)\left[\frac{1}{\lambda_k^2}-\frac{1}{\lambda_{k+1}^2}\right]$$

in which by means of (11)

$$\begin{aligned}\theta_1(\lambda_k x)-\theta_1(\lambda_{k+1}x) &= (\lambda_{k+1}-\lambda_k)x\theta_1'(\bar{x}),\quad \lambda_k x<\bar{x}<\lambda_{k+1}x,\\ &= K\theta(k)x.\end{aligned}$$

Thus

$$\begin{aligned}\sigma_5 &= \sum_{m=s+1,s+3,\cdots}\frac{K\theta(m)}{x^{1/2}\lambda_m^2} + x^{-3/2}\sum_{m=s+1,s+3,\cdots}\theta_1(\lambda_{m+1}x)\left[\frac{1}{\lambda_k^2}-\frac{1}{\lambda_{k+1}^2}\right]\\ &= \frac{K\theta(s+1)}{x^{1/2}\lambda_{s+1}} + \frac{K\theta(s+1)}{x^{3/2}\lambda_{s+1}^2} = \frac{K\theta(n+1)}{n^{1/2}},\qquad \lambda_n x>k_2.\end{aligned}$$

The sum σ_4 remains to be treated. By means of (11) one readily finds

$$\cos(\lambda_m x-\alpha) = \cos[(m\pi+q)x-\alpha] + \frac{K\theta x}{m}$$

and

$$\frac{1}{\lambda_m} - \frac{1}{m\pi} = \frac{K\theta(m)}{m^2}.$$

Hence

$$\left(\frac{2}{\pi x}\right)^{1/2} \sum_{m=s+1}^{\infty} \frac{(-1)^m \cos(\lambda_m x - \alpha)}{\lambda_m}$$
$$= \left(\frac{2}{\pi x}\right)^{1/2} \sum_{m=s+1}^{\infty} \frac{(-1)^m \cos[(m\pi + q)x - \alpha]}{m\pi} + \frac{K\theta(s+1)}{(s+1)^{1/2}}$$

since

$$x\lambda_m > k_2, \quad m > s.$$

The sum on the right is a linear combination of the real and imaginary parts of

$$\left(\frac{2}{\pi x}\right)^{1/2} \sum_{m=s+1}^{\infty} \frac{1}{m\pi}(-1)^m e^{\pi m i x} = \left(\frac{2}{\pi x}\right)^{1/2} \sum_{m=s+1}^{\infty} \frac{1}{m\pi} e^{m\pi i(x-1)}.$$

If $(x-1)=\phi$,

$$S_k(\phi) \equiv S_k = \sum_{m=s+1}^{k} e^{m\pi i\phi} = e^{(s+1)\pi i\phi}\frac{1 - e^{(k-s)\pi i\phi}}{1 - e^{\pi i\phi}} = \frac{K\theta(k, s)}{\phi},$$

then, by the classical transformation of Abel,

$$\sum_{m=s+1}^{\infty} \frac{e^{m\pi i\phi}}{m} = \frac{1}{s+1} S_{s+1} + \frac{1}{s+2}(S_{s+2} - S_{s+1}) + \cdots$$
$$= S_{s+1}\left(\frac{1}{s+1} - \frac{1}{s+2}\right) + S_{s+2}\left(\frac{1}{s+2} - \frac{1}{s+3}\right) + \cdots = \frac{K\theta(\phi, s)}{s\phi}.$$

Therefore

$$\left(\frac{2}{\pi x}\right)^{1/2} \sum_{m=s+1}^{\infty} \frac{(-1)^m e^{m\pi i x}}{m\pi} = \frac{K\theta(s+1, x)}{(1-x)(s+1)^{1/2}}.$$

As has already been pointed out $J_\nu(\lambda_n x)$ is monotone in the sums σ_1 and σ_2, and hence they are readily reduced to sums of terms which alternate in sign and decrease numerically. They will each have the form

$$\frac{K\theta(n)}{n^{1/2}}.$$

The second sum of (13) may be reduced by the methods employed on the first sum. It is found that

$$\sum_{m=n+1}^{\infty} \frac{(-1)^m J_\nu(\lambda_m x)}{\lambda_m} = \frac{K\theta(n, x)}{x^{1/2} n} + \frac{K\theta(n, x)}{x^{3/2} n^{3/2}}.$$

The sum (13) is now readily reduced to the form of Theorem II.

II. The degree of convergence when higher derivatives of $f(x)$ exist

Just as in the case of other well known expansions, the convergence is more rapid when higher derivatives are present provided other conditions are suitably adjusted. The Bessel series requires, in general, rather strong restrictions at the end points of the interval in which the function is represented.

In view of the great similarity of the procedure in Part II to that of Part I, the results in the former will be merely stated and the detailed proofs left to the reader.

Lemma 3. *In the interval* $0 \leqq x \leqq 1$, *let* $f(x)$ *and its first* $(p-2)$ *derivatives be continuous and let* $f^{(p-1)}(x)$ *be absolutely continuous while* $f^{(p)}(x)$ *has bounded variation. Then the coefficient* B_n *of* $J_\nu(\lambda_n x)$ *in the series* (1) *may be written as*

$$B_n = [\pi\lambda_n + K\theta(\lambda_n)] \left\{ \sum_{m=0}^{m=p-1} \sum_{s=0}^{s=p-1} \frac{(-1)^{p+m+s-1} f^{(s)}(1) k(m, s, \nu) J_{\nu+m+1}(\lambda_n)}{\lambda_n^{m+1}} \right.$$

$$\left. + \frac{1}{\lambda_n^p} \int_0^1 J_{\nu+p}(\lambda_n x) \sum_{s=0}^{s=p} \frac{(-1)^s f^{(s)}(x) k(p, s, \nu)}{x^{p-s-1}} dx \right\}$$

where

$$f^{(s)}(x) = \frac{d^s f(x)}{dx^s}$$

and

$k(p, s, \nu)$

$$= \sum_{q=s,1}^{p} \left\{ \left[1 \middle/ \left(1\cdot 3\cdot 5 \cdots \left(q - 2 + \cos^2 \frac{\pi q}{2} \right) \left(\frac{q - 1 - \cos^2 \frac{\pi q}{2}}{2} \right)! \right. \right. \right.$$

$$\left. \left. \times (p - q)!(q - s)!s! \right) \right]$$

$$\times \left[1\cdot 3\cdot 5 \cdots \left(2p - q - 2 + \cos^2 \frac{\pi q}{2} \right) \left(p - \frac{q + 1 + \cos^2 \frac{\pi q}{2}}{2} \right)! \right.$$

$$\left. \left. \times\, q!\nu(\nu + 1) \cdots (\nu + q - s - 1) \right] \right\}$$

in which $\nu(\nu+1)\cdots(\nu+q-s-1)$ *and* $1\cdot3\cdot5\cdots(q-2+\cos^2(\pi q/2))$ *are to be replaced by* 1 *when* $q=s$, $q=1$ *respectively; and the notation* $q=s$, 1 *implies that* $q=s$ *when* $s>0$ *and* $q=1$ *when* $s=0$.

The proof of Lemma 3 is readily obtained after simple but lengthy calculations by integrations by parts with the aid of recurrence formula (8).

LEMMA 4. *Let* $f(x)$ *be a function such as described in Lemma* 3. *Suppose, further, that* $f(x)$ *together with its first* $(p-2)$ *derivatives vanish at the end points* $x=0$ *and* $x=1$. *Then the coefficient* B_n *of the series* (1) *may be written*

$$\frac{(2\pi)^{1/2}\left(\delta_l^2-\cos^2\dfrac{\pi p}{2}\right)f^{(p-1)}(1)(-1)^{n+c}}{\lambda_n^{p-1/2}}+\frac{\pi R(p,\nu)f^{(p-1)}(0)}{\lambda_n^{p}}+\frac{K\theta(n)}{\lambda_n^{p+1/2}}$$

where δ_l *is defined as in Lemma* 2, c *is an undetermined integer and* $R(p,\nu)=\sum_{s=0}^{s=p-1}(-1)^s k(p,s,\nu)$ ($k(p,s,\nu)$ *as in Lemma* 3).

THEOREM III. *Let* $f(x)$ *be a function such as described in Lemma* 4, *and, in addition, let the conditions*

$$\delta_l^2-\cos^2(\pi p/2)f^{(p-1)}(1)=R(p,\nu)f^{(p-1)}(0)=0$$

be satisfied. Then

$$f(x)-S_n(x)=\frac{K\theta(n,x)}{n^{p-1/2}},\qquad 0\leqq x\leqq 1.$$

THEOREM IV. *Let* $f(x)$ *be a function such as described in Lemma* 4. *Then in the interval* $0\leqq a\leqq x\leqq b\leqq 1$

$$f(x)-S_n(x)=\frac{K\theta(x,n)}{x^{1/2}n^p}+\frac{\left(\delta_l^2-\cos^2\dfrac{\pi p}{2}\right)f^{(p-1)}(1)\theta(n,x)}{(1-x)n^{p-1/2}}+\frac{K\theta(n,x)}{x^{3/2}n^{p+1}}$$
$$+\frac{R(p,\nu)f^{(p-1)}(0)K\theta(n,x)}{x^{3/2}n^{p+1/2}}$$

in which $a\neq0$ *unless* $R(p,\nu)f^{(p-1)}(0)=0$, $b\neq1$ *unless*

$$\delta_l^2-\cos^2(\pi p/2)\,f^{(p-1)}(1)=0$$

and

$$\frac{K\theta(n,x)}{x^{1/2}n^p}=\frac{K\theta(n,x)}{x^{3/2}n^{p+1}}\equiv0\qquad\textit{when } a=0.$$

III. On the magnitude of the constants

To make more definite the results of the previous sections, the magnitude of the constants which occur in some of the formulas there used have been computed. Due to the fact that the calculations are rather lengthy and cannot be easily summarized for presentation, these results will be given in an informal manner. It was found for $x \geqq 0$ and $\alpha = (2\nu+1)\pi/4$,

$$J_\nu(x) = \left(\frac{2}{\pi x}\right)^{1/2}\left[\cos(x-\alpha) + \frac{2}{3}\,\frac{\nu^2\theta_1(x)}{x}\right], \qquad (2x)^{1/2} \geqq \nu \geqq 5/2,$$

$$= \left(\frac{2}{\pi x}\right)^{1/2}\left[\cos(x-\alpha) + \frac{\theta_1(x)}{4\cdot 2^{1/2}x}\right], \qquad 0 \leqq \nu \leqq \frac{1}{2^{1/2}}, \quad x > 0,$$

$$= \left(\frac{2}{\pi x}\right)^{1/2}\left[\cos(x-\alpha) + \frac{2^{1/2}\theta_1(x)}{x}\right], \qquad 0 \leqq \nu \leqq 3/2,\ x > 0,$$

$$= \left(\frac{2}{\pi x}\right)^{1/2}\left[\cos(x-\alpha) + \frac{20\theta_1(x)}{3x}\right], \qquad 0 \leqq \nu \leqq 5/2,\ x > 0,$$

that the positive roots of $J_\nu(x)$ larger than

$$\left(\frac{8\nu^2}{3\pi} - \frac{\pi}{2}\right) \text{ when } \nu > 5/2$$

and larger than

$$\left(\frac{80}{3\pi} - \frac{\pi}{2}\right) \text{ when } 0 \leqq \nu \leqq 5/2$$

are given by the formulas

$$\lambda_n = \alpha_1 + (n+k)\pi + \frac{(4/3)\nu^2\theta}{\alpha_1 + \pi(n+k)}, \qquad \nu > 5/2,\ n \geqq 2,$$

$$\lambda_n = \alpha_1 + (n+k)\pi + \frac{20\theta}{3\alpha_1 + 3\pi(n+k)}, \qquad 0 \leqq \nu \leqq 5/2,\ n \geqq 2,$$

$$\alpha_1 = \alpha + \frac{\pi}{2},$$

where k is an integer not less than (-3) for the first equation, and not less than 0 for the second; that the positive roots of $l\lambda J_\nu'(\lambda) + hJ_\nu(\lambda) = 0$ larger than $4\overline{K}/\pi - \pi/2$ are given by the formula

$$\lambda_n = \alpha + (n+k)\pi + \frac{2\overline{K}\theta}{\alpha + \pi(n+k)}, \qquad l \neq 0,\ \alpha = \frac{2\nu+1}{4}\pi,$$

where k is an integer not less than (-3) and

$$\overline{K} = \frac{\pi}{2}\left[\frac{|\nu + h/l|}{1} + \frac{2\cdot 2^{1/2}}{3}(\nu + 1)^2\right], \qquad \nu > 5/2,$$

or k is a positive integer or zero, and

$$\overline{K} = \frac{\pi}{2}\left[2^{1/2} + |\nu + h/l|\right], \qquad 0 \leqq \nu \leqq 5/2.$$

University of Minnesota,
Minneapolis, Minn.

ON THE PROPERTIES OF POLYNOMIALS SATISFYING A LINEAR DIFFERENTIAL EQUATION: PART I*

BY

I. M. SHEFFER

Introduction. Sequences of polynomials such as the Legendre, the Laguerre and the Hermite polynomials appeared in mathematics many years ago, and their properties have been investigated by numerous people. They satisfy simple difference equations, and also are solutions of linear differential equations of second order. The expansion problems (in the complex plane) associated with them are not so old. For the Legendre polynomials the region of convergence was determined by C. Neumann.† More recently the convergence regions for the Laguerre and Hermite polynomials were treated by O. Volk.‡ The paper of Volk considers, more generally, the boundary value problem (in the complex domain) for a second-order linear differential equation, not restricting attention to polynomials. The nth-order equation has since been treated, as a boundary value problem, by L. Bristow.§

Up to the present, however, there has been no general study of the properties of polynomials satisfying a linear differential equation of order higher than two. The present paper has in view such an investigation. There is another aspect to our treatment. In an earlier work we considered the properties of arbitrary *sets* of polynomials,‖ associating with each set a linear differential equation, usually of infinite order. We obtained certain *formal* properties, whose complete justification required convergence proofs. The present paper deals with these matters for the case of a finite order equation.

§1 is preliminary: we state two theorems of Perron, and prove a corollary that is of use later. §2 introduces a fundamental differential equation whose polynomial solutions $\{y_n(x)\}$ we investigate, as well as the entire function

---

* Presented to the Society, December 27, 1929, under the title *The polynomial solutions of linear differential equations*; *Expansions*; received by the editors May 10, 1932.

† *Über die Entwicklung einer Funktion mit imaginärem Argument nach den Kugelfunktionen* 1. *und* 2. *Art*, Halle, 1862.

‡ *Über die Entwicklung von Funktionen einer komplexen Veränderlichen nach Funktionen, die einer linearen Differentialgleichung zweiter Ordnung mit einem Parameter genügen*, Mathematische Annalen, vol. 86 (1922), pp. 296–316.

§ *Expansion theory associated with linear differential equations and their regular singular points*, these Transactions, vol. 33 (1931), pp. 455–474.

‖ *On sets of polynomials and associated linear functional operators and equations*, American Journal of Mathematics, vol. 53 (1931), pp. 15–38. We shall refer to this paper throughout as *Sets*.

solutions $\{\mathcal{D}_n(t)\}$ of the *dual* equation. §3 deals with general theorems on *sets*; §§ 4 and 5 give inequalities and resulting theorems of expansion; and in §6 we obtain certain biorthogonality relations and differential equations for functions allied to $\{y_n(x)\}$ and $\{\mathcal{D}_n(t)\}$.

The important problem of expansions in the polynomials $\{y_n(x)\}$ has hardly been touched. It demands considerations of another order from those of the present paper. Accordingly, we postpone its treatment to Part II.

1. **Preliminary: The Perron theorems.** We have need of the following two theorems (A and B) due to Perron:*

THEOREM A. *Consider the rth-order difference equation*

$$a_{i0}x_i + a_{i1}x_{i+1} + \cdots + x_{i+r} = 0 \qquad (i = 0, 1, \cdots). \tag{i}$$

Let $\lim_{i=\infty} a_{ij}$ *exist,* $=a_j, j=0, 1, \cdots, r-1$, *and let* $q_1, \cdots, q_k$ *be the distinct absolute values of the roots of the characteristic equation*

$$a_0 + a_1 z + \cdots + z^r = 0.$$

Let e_m *= the number of zeros of absolute value* $q_m(e_1 + \cdots + e_m = r)$. *Then if* $a_{i0} \neq 0$ *for all i, there is a fundamental set of r solutions divided into k classes, such that the mth class contains* e_m *of these, and these* e_m *solutions satisfy the condition* $\limsup |x_n|^{1/n} = q_m$.

THEOREM B. *In the system of equations in infinitely many unknowns*

$$\sum_{n=0}^{\infty}(a_n + b_{in})x_{i+n} = c_i \qquad (i = 0, 1, \cdots), \tag{ii}$$

let the following conditions hold:

$$a_0 + b_{i0} \neq 0 \quad (i = 0, 1, \cdots); \quad \limsup |c_i|^{1/i} \leqq 1;$$
$$|b_{in}| \leqq k_i\theta^n, \; 0 < \theta < 1; \quad \lim_{i=\infty} k_i = 0;$$
$$F(z) = \sum_{0}^{\infty} a_n z^n \textit{ is analytic in } |z| \leqq 1.$$

If in $|z| \leqq 1$, $F(z)$ *has n zeros (multiple roots counted multiply), then the general solution of* (ii) *satisfying the condition* $\limsup |x_s|^{1/s} \leqq 1$ *contains n arbitrary constants.*

It is not apparent from the statement of Theorem A that a solution $\{x_n\}$ (not $\equiv 0$) cannot be formed for which $\limsup |x_n|^{1/n} < \min(q_1, \cdots, q_k)$. As we need this fact, we shall establish

* *Über Summengleichungen und Poincarésche Differenzengleichungen*, Mathematische Annalen, vol. 84 (1921), pp. 1–15.

LEMMA 1.* *Under the hypotheses of Theorem* A *there is no solution not identically zero for which* $\limsup |x_n|^{1/n} < \min(q_1, \cdots, q_k)$.

Regarding $x = (x_0, x_1, \cdots)$ as a vector and $L[x]$ as the vector operator that carries x into the vector y with ith component

$$y_i = a_{i0}x_i + a_{i1}x_{i+1} + \cdots + x_{i+r},$$

let us determine an operator M that is inverse to L: $ML[x] \equiv x$. Let the ith component of $M[x]$ be

$$m_{i,i}x_i + m_{i,i+1}x_{i+1} + \cdots + m_{i,i+n}x_{i+n} + \cdots .$$

Then we are to have, identically in the $\{x_i\}$,

$$\sum_{i=s}^{\infty} m_{si}(a_{i0}x_i + a_{i1}x_{i+1} + \cdots + x_{i+r}) = x_s.$$

This gives the equations

$$\begin{aligned} &m_{ss}a_{s0} = 1, \\ \text{(a)} \qquad &m_{ss}a_{si} + m_{s,s+1}a_{s+1,i-1} + \cdots + m_{s,s+i}a_{s+i,0} = 0 \qquad (i = 1, \cdots, r); \\ &m_{s,s+j}a_{s+j,r} + m_{s,s+j+1}a_{s+j+1,r-1} + \cdots + m_{s,s+j+r}a_{s+j+r,0} = 0 \\ &\qquad\qquad (j = 1, 2, \cdots). \end{aligned}$$

Since $a_{i0} \neq 0$ for all i, the quantities m_{ij} exist and are unique. M is then determined. It remains to consider convergence. Let *s be fixed*, and set

$$\text{(b)} \qquad n_k = m_{s,s+k}, \quad b_{k,r-i} = a_{s+k,i}.$$

Then we have the equations

$$\text{(c)} \qquad b_{i0}n_i + b_{i+1,1}n_{i+1} + \cdots + b_{i+r,r}n_{i+r} = 0 \qquad (i = 0, 1, \cdots)$$

for $n_0, n_1, \cdots$. It is easily verified that the conditions of Theorem A hold for (c), so that for every solution $\{n_k\}$ of (c) we have $\limsup |n_k|^{1/k} \leqq \max(|t_1|, \cdots, |t_r|)$, where $t_1, \cdots, t_r$ are the zeros of $1 + a_{r-1}t + \cdots + a_0t^r = 0$. Now if $\min(q_1, \cdots, q_k) = 0$ the lemma is vacuously true. We may then assume that $\min(q_1, \cdots, q_k) = \lambda > 0$, in which case $a_0 \neq 0$. Then $t_1, \cdots, t_r$ are the reciprocals of the roots of the characteristic equation of (i), so that $\limsup |n_k|^{1/k} \leqq 1/\lambda$. To $\epsilon > 0$ we have $|m_{s,s+k}| \leqq K_s(\epsilon)/(\lambda - \epsilon)^k$ for all s.

Now suppose a solution $\{x_n\}$ of (i) exists such that $\limsup |x_n|^{1/n} = \delta < \lambda$. Then to $\epsilon' > 0$ we have $|x_n| < C(\epsilon')(\delta + \epsilon')^n$. Let $\gamma = \max(|a_{i0}|, \cdots, |a_{ir}|)$ for all i. Then

* A statement, without proof, of this lemma is given in Nörlund, *Differenzenrechnung*, 1924, p. 309.

$$| \{L[x]\}_i | = | a_{i0}x_i + \cdots + x_{i+r} |$$
$$\leqq C\gamma[(\delta + \epsilon')^i + \cdots + (\delta + \epsilon')^{i+r}] = C\gamma H(\delta + \epsilon')^i,$$

where the definition of H is obvious. Therefore

$$| \{M[L[x]]\}_s | \leqq K_s CH\gamma\left[\frac{(\delta + \epsilon')^s}{(\lambda - \epsilon)^0} + \frac{(\delta + \epsilon')^{s+1}}{(\lambda - \epsilon)} + \cdots\right]$$
$$= K_s CH\gamma(\delta + \epsilon')^s \sum_{t=0}^{\infty} \left(\frac{\delta + \epsilon'}{\lambda - \epsilon}\right)^t.$$

On choosing ϵ, ϵ' small enough the infinite geometric series converges. Hence, when we substitute $L[x]$ into M, forming $ML[x]$, and in the sth component combine coefficients of the same x_i's, we obtain an absolutely convergent series; the process is then legitimate. But $L[x] = (0, 0, \cdots)$, and since $ML[x] \equiv x$, it follows that $x = (0, 0, \cdots)$. This proves the lemma.

2. **Solutions of a differential equation and its dual.** Our principal aim is the study of the polynomial solutions of the kth-order linear differential equation

$$L[y(x)] \equiv L_0(x)y(x) + L_1(x)y'(x) + \cdots + L_k(x)y^{(k)}(x) = \lambda y(x), \tag{1}$$

where

$$L_i(x) = l_{i0} + l_{i1}x + \cdots + l_{ii}x^i \qquad (i = 0, 1, \cdots, k) \tag{2}$$

is a polynomial of degree not exceeding i and λ is a parameter, and of its dual equation (soon to be defined). We *define* λ_n by

$$\lambda_n = l_{00} + nl_{11} + n(n-1)l_{22} + \cdots + n(n-1)\cdots(n-k+1)l_{kk}. \tag{3}$$

THEOREM 1. *If* $\lambda_m \neq \lambda_n$, $m \neq n$, and if $l_{kk} \neq 0$, the equation*

$$L[y(x)] = \lambda y(x) \tag{4}$$

has an entire function solution ($\not\equiv 0$) if and only if λ has one of the values $\lambda = \lambda_0, \lambda_1, \cdots$; and when $\lambda = \lambda_n$ there is just one entire function solution, namely a polynomial $y_n(x)$ of degree exactly n.

To demonstrate this, substitute into (2) the power series $y(x) = \sum_0^\infty y_n x^n$. On equating coefficients we find the following equations for the y_i:

$$(\lambda_n - \lambda)y_n + \sigma_{n,n+1}y_{n+1} + \sigma_{n,n+2}y_{n+2} + \cdots + \sigma_{n,n+k}y_{n+k} = 0 \tag{5}$$
$$(n = 0, 1, \cdots),$$

* If $l_{kk} = 0$ or $\lambda_m = \lambda_n$ for some $m \neq n$, it is necessary to modify some of our later arguments, and we leave such considerations out of the present paper.

where

$$
\begin{aligned}
\sigma_{n,n+1} &= l_{10}(n+1) + l_{21}(n+1)n + \cdots + l_{k,k-1}(n+1)n\cdots(n-k+2),\\
\sigma_{n,n+2} &= l_{20}(n+2)(n+1) + l_{31}(n+2)(n+1)n + \cdots\\
&\quad + l_{k,k-2}(n+2)(n+1)\cdots(n-k+3),\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
\sigma_{n,n+k} &= l_{k0}(n+k)(n+k-1)\cdots(n+1).
\end{aligned}
\tag{6}
$$

For definiteness let us suppose that* $l_{k0}\neq 0$. We then obtain the difference equation

$$
\frac{\lambda_n-\lambda}{\sigma_{n,n+k}}y_n + \frac{\sigma_{n,n+1}}{\sigma_{n,n+k}}y_{n+1} + \cdots + y_{n+k} = 0 \quad (n = 0, 1, \cdots), \tag{5'}
$$

with

$$
\lim_{n=\infty}\frac{\lambda_n-\lambda}{\sigma_{n,n+k}} = \frac{l_{kk}}{l_{k0}}, \quad \lim_{n=\infty}\frac{\sigma_{n,n+i}}{\sigma_{n,n+k}} = \frac{l_{k,k-i}}{l_{k0}} \quad (i = 1, \cdots, k-1). \tag{7}
$$

The characteristic equation of (5′) is

$$
l_{kk} + l_{k,k-1}t + \cdots + l_{k0}t^k = 0 \qquad (l_{kk}\neq 0). \tag{8}
$$

Let $\alpha(>0)$ be the least absolute value of the roots of (5′). Then, by Lemma 1, for *every* solution $\{y_n\}$ we have $\limsup |y_n|^{1/n} \geqq \alpha$ provided $\lambda\neq\lambda_0, \lambda_1, \cdots$. Hence in this case $y(x) = \sum_0^\infty y_n x^n$, with radius of convergence $\leqq 1/\alpha$, cannot be an entire function (unless $y(x)\equiv 0$).

Now let $\lambda = \lambda_n$. A solution of equations (5) is seen to be $y_{n+1} = y_{n+2} = \cdots = 0$, y_n arbitrary (but $\neq 0$), and $y_{n-1}, y_{n-2}, \cdots, y_0$ determined uniquely and successively (in view of $\lambda_m\neq\lambda_n$) from the $(n-1)$st, $(n-2)$d, $\cdots$, 0th equations of (5). Hence one entire function solution of (4) for($\lambda=\lambda_n$) is the polynomial

$$
y_n(x) = y_{n0} + y_{n1}x + \cdots + y_{nn}x^n, \quad y_{nn}\neq 0. \tag{9}
$$

To show that there is no other entire function solution, ignore the first $n+1$ equations of (5′). The system of equations remaining has the limits (7) for its coefficients. Moreover, *now* the coefficients of that y_s of lowest index in each equation is different from zero, so that Lemma 1 again applies, and the only entire function solution (for the modified system of equations) is the function *zero*. That is, the only entire function solution of (4) is one whose coefficients beyond x^n are all zero. And, since a polynomial solution of degree $\leqq n$ is unique, as we have seen, the theorem is established.

* If $l_{k0}=0$, (5) becomes a difference equation of order $<k$, but the same conclusion will follow.

DEFINITION. *The numbers* $\{\lambda_n\}$ *may be defined as the characteristic numbers of equation* (4).

Let us in (1) replace each $y^{(i)}$ by t^i, thus obtaining the function

$$L(t; x) = L_0(x) + L_1(x)t + \cdots + L_k(x)t^k; \tag{10}$$

or,

$$\begin{aligned} L(t; x) &= (l_{00} + l_{10}t + \cdots + l_{k0}t^k) + x(l_{11}t + l_{21}t^2 + \cdots \\ &\qquad + l_{k1}t^k) + \cdots + x^k(l_{kk}t^k) \\ &= \mathcal{L}_0(t) + \mathcal{L}_1(t)x + \cdots + \mathcal{L}_k(t)x^k. \end{aligned} \tag{10'}$$

Now in (10′) replace each power x^i by the derivative $\mathcal{D}^{(i)}(t)$ of a function* $\mathcal{D}(t)$. This gives us a differential expression which we term the *dual* of (1):

$$\mathcal{L}[\mathcal{D}(t)] \equiv \mathcal{L}_0(t)\mathcal{D}(t) + \mathcal{L}_1(t)\mathcal{D}'(t) + \cdots + \mathcal{L}_k(t)\mathcal{D}^{(k)}(t). \tag{11}$$

Evidently, (1) is the *dual* of (11). Similarly, we term equations (4) and

$$\mathcal{L}[\mathcal{D}(t)] \equiv \lambda\mathcal{D}(t) \tag{12}$$

dual equations.

THEOREM 2. *If* $\lambda_m \neq \lambda_n$, $m \neq n$, *and* $l_{kk} \neq 0$, *equation* (12) *has a formal power series solution about the origin if and only*† *if* $\lambda = \lambda_0, \lambda_1, \cdots$; *and when* $\lambda = \lambda_n$ *there is precisely one such power series,* $\mathcal{D}_n(t)(\not\equiv 0)$, *and it is an entire function with an nth order zero at the origin.*

To prove this substitute the power series $\mathcal{D}(t) = \sum d_n t^n$ into (12) and equate coefficients. This yields the equations

$$(\lambda_n - \lambda)d_n + \alpha_{n,n-1}d_{n-1} + \alpha_{n,n-2}d_{n-2} + \cdots + \alpha_{n,n-k}d_{n-k} = 0 \quad (n = 0, 1, \cdots) \tag{13}$$

where

$$\begin{aligned} \alpha_{n,n-1} &= l_{10} + l_{21}(n-1) + l_{32}(n-1)(n-2) + \cdots \\ &\qquad + l_{k,k-1}(n-1)(n-2)\cdots(n-k+1), \\ \alpha_{n,n-2} &= l_{20} + l_{31}(n-2) + \cdots + l_{k,k-2}(n-2)\cdots(n-k+1), \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \alpha_{n,n-k} &= l_{k0}. \end{aligned} \tag{14}$$

If $\lambda \neq \lambda_0, \lambda_1, \cdots$ then successively we get $d_0 = d_1 = \cdots = 0$, so that the only formal power series solution is $\mathcal{D}(t) \equiv 0$. Now suppose $\lambda = \lambda_n$. Then $d_0 = d_1 = \cdots = d_{n-1} = 0$, d_n is arbitrary, and $d_{n+1}, d_{n+2}, \cdots$ are successively

* See *Sets*, loc. cit., pp. 31–32.

† Consequently, we may say that equations (4) and (12) have the same characteristic numbers.

and uniquely determined (in terms of d_n). Hence there is just one formal solution, $\sum_{i=n}^{\infty} d_i t^i$. We proceed to show this solution is an entire function. The first $n+1$ equations of (13) drop out, giving us

$$(\lambda_s - \lambda_n)d_s + \alpha_{s,s-1}d_{s-1} + \cdots + \alpha_{s,s-k}d_{s-k} = 0 \qquad (s = n+1, n+2, \cdots);$$

and, on setting $r = s-k$, we get the difference equation

$$(13') \qquad d_{r+k} + \frac{\alpha_{r+k,r+k-1}}{\lambda_{r+k} - \lambda_n} d_{r+k-1} + \cdots + \frac{\alpha_{r+k,r}}{\lambda_{r+k} - \lambda_n} d_r = 0$$
$$(r = n+1-k, n+2-k, \cdots),$$

with

$$(15) \qquad \lim_{r=\infty} \frac{\alpha_{r+k,r+k-i}}{\lambda_{r+k} - \lambda_n} = 0 \qquad (i = 1, \cdots, k),$$

and with the characteristic equation

$$(16) \qquad t^k = 0.$$

By* the Perron Theorem A, for every solution of (13′) we have $\limsup |d_r|^{1/r} = 0$; and this implies that $\mathcal{D}(t)$ is an entire function.

We can say even more:

COROLLARY. *The solution* $\mathcal{D}_n(t) = \sum_{i=n}^{\infty} d_{ni} t^i$ *corresponding to* $\lambda = \lambda_n$ *satisfies the inequality*

$$\limsup_{s=\infty} |\mathcal{D}_n^{(s)}(0)|^{1/s} \leqq \rho = \textit{maximum absolute value of the zeros of } L_k(x),$$

so that the function $\Delta_n(t)$ *defined by*

$$(17) \qquad \Delta_n(t) = \sum_{i=n}^{\infty} i!\, d_{ni} t^i$$

has a radius of convergence at least equal to $1/\rho$.

To show this, let $d_{ni} = v_i/i!$; then (13′) leads to the difference equation

$$(13'') \qquad \begin{aligned} v_{r+k} &+ \frac{\alpha_{r+k,r+k-1}}{\lambda_{r+k} - \lambda_n}(r+k)v_{r+k-1} + \cdots \\ &+ \frac{\alpha_{r+k,r}}{\lambda_{r+k} - \lambda_n}(r+k)(r+k-1)\cdots(r+1)v_r = 0, \end{aligned}$$

with

* If $l_{k0} = 0$, (13′) is a difference equation of order less than k, but the same conclusion follows.

$$\lim_{r=\infty} \frac{\alpha_{r+k,r+k-i}}{\lambda_{r+k} - \lambda_n}(r + k)(r + k - 1) \cdots (r + k - i + 1) = \frac{l_{k,k-i}}{l_{kk}}$$
$$(i = 1, \cdots, k),$$

and to the characteristic equation $L_k(t) = 0$, whose largest zero is in absolute value ρ. Now the coefficient of v_r in (13″) is never zero.* Hence, from Theorem A, $\limsup |v_r|^{1/r} \leqq \rho$; and from this the corollary follows.

$\mathcal{D}_n(t)$ is the so-called Borel entire function associated with $\Delta_n(t)$ and the two are related by the following integral:

$$\mathcal{D}_n(t) = \frac{1}{2\pi i} \int_C \frac{e^{tu}}{u} \Delta_n\left(\frac{1}{u}\right) du, \tag{18}$$

where C is a closed contour surrounding the origin and lying wholly outside of $|u| = \rho$.

Similarly, if we set

$$Y_n(x) = 0!\, y_{n0} + 1!\, y_{n1} x + \cdots + n!\, y_{nn} x^n, \tag{19}$$

then $y_n(x)$ is the Borel entire function for $Y_n(x)$, and we have

$$y_n(x) = \frac{1}{2\pi i} \int_\Gamma \frac{e^{xu}}{u} Y_n\left(\frac{1}{u}\right) du, \tag{20}$$

Γ being a closed contour surrounding the origin.

DEFINITION. *An analytic function $f(t; x)$ is self-dual with respect to the above operators L, $\mathcal{L}$, operating respectively on the variables x and t, if*

$$L[f(t; x)] = \mathcal{L}[f(t; x)].$$

COROLLARY. *e^{tz} is a self-dual, and*

$$L[e^{tz}] = \mathcal{L}[e^{tz}] = e^{tz} L(t; x). \tag{21}$$

3. Associated sets of functions; the sets P_λ, Q_λ. Let us now consider the following parametric differential expressions corresponding to (1) and (11):

$$L_\lambda[y(x)] \equiv L[y(x)] - \lambda y(x), \tag{22}$$

$$\mathcal{L}_\lambda[\mathcal{D}(t)] \equiv \mathcal{L}[\mathcal{D}(t)] - \lambda \mathcal{D}(t). \tag{23}$$

Define the *set* of polynomials P_λ: $\{P_n(x; \lambda)\}$ by

$$\begin{aligned} P_n(x; \lambda) = L_\lambda[x^n] &= (L_0(x) - \lambda)x^n + nL_1(x)x^{n-1} + \cdots \\ &\quad + n(n-1) \cdots (n - k + 1)L_k(x)x^{n-k} \quad (n = 0, 1, \cdots), \end{aligned} \tag{24}$$

---

* That is, if $l_{k0} \neq 0$. Should $l_{k0} = 0$ (here and hereafter), the remark of a previous footnote applies.

and the *set* of functions $\mathcal{P}_\lambda$: $\{\mathcal{P}_n(t;\lambda)\}$ by

$$\text{(25)}\quad \begin{aligned}\mathcal{P}_n(t;\lambda) = \mathcal{L}_\lambda[t^n] = (\mathcal{L}_0(t) - \lambda)t^n + n\mathcal{L}_1(t)t^{n-1} + \cdots \\ + n(n-1)\cdots(n-k+1)\mathcal{L}_k(t)t^{n-k} \quad (n = 0, 1, \cdots).\end{aligned}$$

We see that $P_n(x;\lambda)$ is a polynomial in x, of degree n if $\lambda\neq\lambda_0, \lambda_1, \cdots$, and that $\mathcal{P}_n(t;\lambda)$ is a polynomial in t, of degree not exceeding $n+k$, with a zero at the origin of order at least n.

Definition. *Let $H(t;x)$ be a symbol for a formal power series in t with coefficients that are formal power series in x, so that when it is expressed formally as a power series in x, the coefficients are (formal) power series in t:*

$$H(t;x) \sim \sum_0^\infty h_n(x)t^n/n! \sim \sum_0^\infty \mathcal{H}_n(t)x^n/n! \qquad (h_n(x),\ \mathcal{H}_n(t) \textit{ power series}).$$

Then we say that the two sets of functions $\{h_n(x)\}$, $\{\mathcal{H}_n(t)\}$ $(n=0, 1, \cdots)$ are associated sets.

Lemma 2. *The two sets P_λ, $\mathcal{P}_\lambda$ given in* (24, 25) *are associated sets, and, for* all x and t,*

$$\text{(26)}\quad \sum_{n=0}^\infty P_n(x;\lambda)t^n/n! = \sum_{n=0}^\infty \mathcal{P}_n(t;\lambda)x^n/n! = e^{tx}L_\lambda(t;x) = (\textit{definition})P(t,x;\lambda),$$

where

$$\text{(27)}\qquad L_\lambda(t;x) \equiv L(t;x) - \lambda.$$

(26) follows from (24), (25) and (21). The convergence of (26) is immediate.

If we expand (24), we obtain

$$\text{(28)}\quad P_n(x;\lambda) = (\lambda_n - \lambda)x^n + \sigma_{n-1,n}x^{n-1} + \sigma_{n-2,n}x^{n-2} + \cdots + \sigma_{n-k,n}x^{n-k},$$

where the σ_{ij} are given by (6).

In our theory of sets of polynomials† we considered the *multiplication* of sets. Thus, if P: $\{P_n(x)\}$, Q: $\{Q_n(x)\}$ are any two sets, where

$$P_n(x) = p_{n0} + p_{n1}x + \cdots + p_{nn}x^n,\ Q_n(x) = q_{n0} + q_{n1}x + \cdots + q_{nn}x^n,$$

then PQ is the set $\{PQ_n(x)\}$, where

$$PQ_n(x) = p_{n0}Q_0(x) + p_{n1}Q_1(x) + \cdots + p_{nn}Q_n(x) \quad (n = 0, 1, \cdots).$$

In particular, a set Q is the *inverse* of P if $PQ=I$ where I is the identity set: $I_n(x)=x^n$. It is easy to see that an inverse Q exists if and only if $P_n(x)$ is of

* (26) is formally true in the general theory of sets of polynomials.

† *Sets*, p. 16.

degree exactly n for every n; and in this case Q is unique, and P is also the inverse of Q: $QP=I$.

Now $P_n(x;\lambda)$ is of degree exactly $n(\lambda\neq\lambda_0, \lambda_1, \cdots)$. Hence the set P_λ possesses an *inverse* set Q_λ: $\{Q_n(x;\lambda)\}$. Since $P_\lambda Q_\lambda=I$, there follows from (28)

LEMMA 3. *The set* $\{Q_n(x;\lambda)\}$ *satisfies the difference equation*

$$(\lambda_n-\lambda)Q_n(x;\lambda)+\sigma_{n-1,n}Q_{n-1}(x;\lambda)+\cdots+\sigma_{n-k,n}Q_{n-k}(x;\lambda)=x^n \quad (n=0,1,\cdots). \tag{29}$$

Examination of the first few $Q_n(x;\lambda)$'s suggests

LEMMA 4. $Q_n(x;\lambda)$ *is a rational function of* λ, *having at most* simple poles at* $\lambda=\lambda_0, \lambda_1, \cdots, \lambda_n$.

This is true for $n=0$, 1 as is readily seen. The Lemma then follows by induction from (29).

From (24) we find that

$$L_\lambda[Q_n(x;\lambda)]=QP_n(x;\lambda)=x^n \qquad (n=0,1,\cdots). \tag{30}$$

If we multiply through by $t^n/n!$ and sum formally from $n=0$ to ∞, we obtain

$$L_\lambda[Q(t,x;\lambda)]=e^{tx} \tag{31}$$

where

$$Q(t,x;\lambda)=\sum_{n=0}^{\infty}Q_n(x;\lambda)t^n/n!. \tag{32}$$

We proceed to show that† (32) is uniformly convergent in t and x in some region, thus making (31) valid.

On dividing (29) through by $\sigma_{n-k,n}$, we obtain a system of equations for $Q_0, Q_1, \cdots$ satisfying all the hypotheses of Perron Theorem B, the function $F(z)$ here being $L_k(z)/l_{k0}$, provided $|x|\leqq 1$. If $L_k(z)$ does not have all its k zeros in $|z|\leqq 1$, we cannot apply Theorem B to all the solutions of the system in question. This difficulty can be overcome by modifying the system as follows:

Let ρ again denote the largest absolute value of the zeros of $L_k(z)$ and set $Q_s(x;\lambda)=\rho^s R_s(x;\lambda)$. We then obtain for $\{R_s\}$ the system of equations

---

* Individual $Q_n(x;\lambda)$'s may fail to have a pole at some of these points; e.g., $Q_1(x;\lambda)$ will not have λ_0 as a pole if $l_{10}=0$. It is however easy to prove that $Q_n(x;\lambda)$ always has λ_n as a pole (i.e., if x is not given special values).

† $Q(t,x;\lambda)$ is a function of λ as well; we must therefore avoid such values of λ as will make $Q(t,x;\lambda)$ singular.

$$(\alpha)\quad R_s+\frac{\sigma_{k+s,s+1}}{\sigma_{s,s+k}}\rho R_{s+1}+\cdots+\frac{\lambda_{k+s}-\lambda}{\sigma_{s,s+k}}\rho^k R_{s+k}=\frac{x^{s+k}}{\sigma_{s,s+k}\rho^s}\quad(s=0,1,\cdots).$$

This system satisfies the conditions of Theorem B, if we restrict x to lie in $|x|\leqq\rho$, and the function $F(z)$ is here $(1/l_{k0})L_k(z\rho)$. Now the characteristic roots all lie in $|z|\leqq 1$ so that by Theorem B, the general solution of (α) for which $\lim\sup|R_s|^{1/s}\leqq 1$ contains k arbitrary constants, exactly as many as enter into the general solution of (α). Hence, *every* solution $\{R_s\}$ of (α) satisfies the condition $\lim\sup|R_s|^{1/s}\leqq 1$. We thus have

COROLLARY 1. *For* $|x|\leqq\rho$,

$$(33)\qquad \limsup_{n=\infty}|Q_n(x;\lambda)|^{1/n}\leqq\rho\qquad(\lambda\neq\lambda_0,\lambda_1,\cdots,\lambda_n).$$

We have, moreover, *uniformly** in $|x|\leqq\rho$,

$$(33')\qquad |Q_n(x;\lambda)|\leqq C(\rho+\epsilon)^n.$$

Here $\epsilon>0$ is arbitrary, and C does not depend on x.

If we had defined $R_s(x;\lambda)$ by $Q_s(x;\lambda)=\delta^s R_s(x;\lambda)$, $\delta\geqq\rho$, the argument made above would continue to hold, giving us

COROLLARY 2. *For* $|x|\leqq\delta$, $\delta(\geqq\rho)$ *arbitrary,*

$$(34)\qquad \limsup_{n=\infty}|Q_n(x;\lambda)|^{1/n}\leqq\delta\qquad(\lambda\neq\lambda_0,\lambda_1,\cdots,\lambda_n);$$

and

$$(34')\qquad |Q_n(x;\lambda)|\leqq C_\delta(\delta+\epsilon)^n$$

uniformly in $|x|\leqq\delta$. *Here* $\epsilon>0$ *is arbitrary, and* C_δ *depends only on* ϵ *and* δ.

Since δ may be chosen arbitrarily large we have

THEOREM 3. *The function* $Q(t,x;\lambda)$ *given by* (32) *is analytic*† *in* t, x, λ; *it is an entire function in* t *and* x, *and its singularities in* λ *are at most at the points* $\lambda=\lambda_0,\lambda_1,\cdots$. *Moreover* (31) *holds for every* t, x, λ $(\lambda\neq\lambda_0,\lambda_1,\cdots)$.

Let us return to the theory of *sets*.‡ We have there given the

DEFINITION. *A triangular set of functions* $\{\mathfrak{P}_n(t)\}$, $n=0,1,\cdots$, *is a set of formal power series in* t *such that* $\mathfrak{P}_n(t)$ *begins with a power of* t *not less than* n:

$$\mathfrak{P}_n(t)\sim\pi_{nn}t^n+\pi_{n,n+1}t^{n+1}+\cdots.$$

---

* The uniformity can be established from the Perron proofs.

† Relation (34′) is also uniform in every bounded λ-region, the points $\lambda=\lambda_0,\lambda_1,\cdots$ being deleted.

‡ *Sets*, p. 32.

For such sets multiplication is defined as follows:

$$\mathfrak{P}\mathfrak{Q} \equiv \{\mathfrak{P}\mathfrak{Q}_n(t)\}: \mathfrak{P}\mathfrak{Q}_n(t) \sim \pi_{nn}\mathfrak{Q}_n(t) + \pi_{n,n+1}\mathfrak{Q}_{n+1}(t) + \cdots \quad (n = 0, 1, \cdots).$$

If, in particular, $\mathfrak{P}\mathfrak{Q} = \mathfrak{I}$ where $\mathfrak{I}$: $\{\mathfrak{I}_n(t) = t^n\}$ is the *identity* set, then $\mathfrak{Q}$ is the *inverse* of $\mathfrak{P}$. Such $\mathfrak{Q}$ exists if and only if $\pi_{nn} \neq 0$, $n = 0, 1, \cdots$, and then $\mathfrak{Q}$ is unique, and $\mathfrak{P}$ is the inverse of $\mathfrak{Q}$.

Let P: $\{P_n(x)\}$ be a polynomial set, and $\mathfrak{P}$: $\{\mathfrak{P}_n(t)\}$ the associated set. ($\mathfrak{P}$ is then a triangular set.) On setting

$$P_n(x) = p_{n0} + p_{n1}x + \cdots + p_{nn}x^n, \quad \mathfrak{P}_n(t) \sim \pi_{nn}t^n + \pi_{n,n+1}t^{n+1} + \cdots,$$

it is seen that the property of being associated sets is equivalent to the relations

$$(35) \qquad \begin{aligned} \pi_{nn} &= p_{nn}, \\ \pi_{n,n+i} &= p_{n+i,n}/((n+i)(n+i-1)\cdots(n+1)) \quad (i = 1, 2, \cdots). \end{aligned}$$

We can establish the following general theorem on sets:

THEOREM 4. (a) *Let P, $\mathfrak{P}$ be associated sets, and Q, $\mathfrak{Q}$ their respective inverses. Then Q, $\mathfrak{Q}$ are associated sets.*

(b) *If P, Q are inverse sets, and $\mathfrak{P}$, $\mathfrak{Q}$ their associated sets, then $\mathfrak{P}$, $\mathfrak{Q}$ are inverse sets.*

Consider (a). Let

$$Q_n(x) = q_{n0} + \cdots + q_{nn}x^n, \qquad \mathfrak{Q}_n(t) \sim \kappa_{nn}t^n + \kappa_{n,n+1}t^{n+1} + \cdots.$$

From $PQ = I$ and $\mathfrak{P}\mathfrak{Q} = \mathfrak{I}$ we obtain the relations

$$p_{n0}Q_0(x) + \cdots + p_{nn}Q_n(x) = x^n, \quad \pi_{nn}\mathfrak{Q}_n(t) + \pi_{n,n+1}\mathfrak{Q}_{n+1}(t) + \cdots \sim t^n,$$

$$(\alpha) \quad p_{nn}q_{n,n-i} + p_{n,n-1}q_{n-1,n-i} + \cdots + p_{n,n-i}q_{n-i,n-i} = \begin{cases} 1, & i = 0, \\ 0, & i = 1, \cdots, n; \end{cases}$$

$$(\beta) \quad \pi_{nn}\kappa_{n,n+i} + \pi_{n,n+1}\kappa_{n+1,n+i} + \cdots + \pi_{n,n+i}\kappa_{n+i,n+i} = \begin{cases} 1, & i = 0, \\ 0, & i = 1, 2, \cdots. \end{cases}$$

The theorem will be proved if we establish that

$$(\mathrm{i}) \qquad \begin{aligned} \kappa_{nn} &= q_{nn}, \\ \kappa_{n,n+i} &= q_{n+i,n}/((n+i)\cdots(n+1)) \qquad (i = 1, 2, \cdots), \end{aligned}$$

or, that $s_{n,n+i} = \kappa_{n,n+i}$, $i = 0, 1, \cdots$, where

$$(\mathrm{ii}) \qquad \begin{aligned} s_{nn} &= q_{nn}, \\ s_{n,n+i} &= q_{n+i,n}/((n+i)\cdots(n+1)). \end{aligned}$$

In equations (α), substitute for $q_{n,n+i}$ its value in terms of $s_{n+i,n}$. Moreover, in the resulting equations, leave the first equation unchanged, replace n by

$n+1$ in the second, n by $n+2$ in the third, and so on. This gives the following equations (α'), equivalent to equations (α):

$$(\alpha') \quad \pi_{n+i,n+i}s_{n,n+i} + \pi_{n+i-1,n+i}s_{n,n+i-1} + \cdots + \pi_{n,n+i}s_{nn} = \begin{cases} 1, & i = 0, \\ 0, & i = 1, 2, \cdots. \end{cases}$$

We verify at once that $s_{nn} = \kappa_{nn}$, $s_{n,n+1} = \kappa_{n,n+1}$. Suppose $s_{n,n+j} = \kappa_{n,n+j}$ for $j = 0, 1, \cdots, i-1$. We shall then prove it for $j = i$, and the theorem will be demonstrated. Denote respectively by $E_{n0}, E_{n1}, \cdots; F_{n0}, F_{n1}, \cdots$ the left hand members of (α') and (β) for $i = 0, 1, \cdots$. Now form linear expressions in the s's and κ's, respectively:

$$E_i \equiv \pi_{n,n+i}E_{n+i,0} + \pi_{n,n+i-1}E_{n+i-1,1} + \cdots + \pi_{nn}E_{ni} = \pi_{n,n+i},$$

$$F_i \equiv \pi_{n,n+i}F_{n0} + \pi_{n+1,n+i}F_{n1} + \cdots + \pi_{n+i,n+i}F_{ni} = \pi_{n,n+i}.$$

The coefficient in E_i of $s_{n+i-p,n+i-p+q}$ is readily seen to be $\pi_{n,n+i-p}\,\pi_{n+i-p+q,n+i}$, and this is also the coefficient of $\kappa_{n+i-p,n+i-p+q}$ in F_i. Hence from $E_i = F_i$, and our induction assumption, it follows that $\pi_{nn}\,\pi_{n+i,n+i}\,s_{n,n+i} = \pi_{n+i,n+i}\pi_{nn}\kappa_{n,n+i}$. But $\pi_{nn} = p_{nn} \neq 0$, $n = 0, 1, \cdots$ (since P possesses an inverse). Therefore, $s_{n,n+i} = \kappa_{n,n+i}$ and the induction is complete.

To establish (b), let $\mathfrak{Q}^*$ be the inverse of $\mathfrak{P}$. Then by (a), Q and $\mathfrak{Q}^*$ are associated sets. But the associate of a set is unique, so that $\mathfrak{Q} \equiv \mathfrak{Q}^*$.

Consider the associate sets P_λ, $\mathfrak{P}_\lambda$ of (26). The coefficient of t^n in $\mathfrak{P}_n(t; \lambda)$ is not zero for $\lambda \neq \lambda_0, \lambda_1, \cdots$. Hence $\mathfrak{P}_\lambda$ possesses an inverse $\mathfrak{Q}_\lambda$: $\{\mathfrak{Q}_n(t; \lambda)\}$, and by Theorem 4 we have the

COROLLARY. *Q_λ and $\mathfrak{Q}_\lambda$ are associated sets, and*

$$(36) \qquad Q(t, x; \lambda) = \sum_{n=0}^{\infty} Q_n(x; \lambda)t^n/n! = \sum_{n=0}^{\infty} \mathfrak{Q}_n(t; \lambda)x^n/n!,$$

the series converging uniformly in every bounded x, t, λ region (on deleting the points $\lambda_0, \lambda_1, \cdots$).

From the definition (25) of $\mathfrak{P}_\lambda$ we have (using (14))

$$(37) \quad \mathfrak{P}_n(t; \lambda) = (\lambda_n - \lambda)t^n + \alpha_{n+1,n}t^{n+1} + \alpha_{n+2,n}t^{n+2} + \cdots + \alpha_{n+k,n}t^{n+k}.$$

On equating corresponding coefficients in $\mathfrak{P}_\lambda\mathfrak{Q}_\lambda = \mathfrak{J}$ there results the following difference equation for $\mathfrak{Q}_n(t; \lambda)$:

$$(38) \quad \begin{aligned} &(\lambda_n - \lambda)\mathfrak{Q}_n(t; \lambda) + \alpha_{n+1,n}\mathfrak{Q}_{n+1}(t; \lambda) + \cdots + \alpha_{n+k,n}\mathfrak{Q}_{n+k}(t; \lambda) = t^n \\ &\qquad\qquad (n = 0, 1, \cdots). \end{aligned}$$

Now $\mathfrak{Q}_\lambda$ is the inverse of $\mathfrak{P}_\lambda$, so that we have

$$(39) \qquad \mathfrak{L}_\lambda[\mathfrak{Q}_n(t; \lambda)] = t^n,$$

$$\mathcal{L}_\lambda[Q(t, x; \lambda)] = L_\lambda[Q(t, x; \lambda)] = e^{tx}. \quad \text{(See (31, 36).)} \tag{40}$$

On applying the operator $\mathcal{L}_\lambda$ to (32), we then have $e^{tx} = \sum_0^\infty Q_n(x; \lambda)\mathcal{L}_\lambda[t^n]/n!$, or

$$e^{tx} = \sum_{n=0}^{\infty} Q_n(x; \lambda)\mathcal{P}_n(t; \lambda)/n!. \tag{41}$$

Similarly we obtain

$$e^{tx} = \sum_{n=0}^{\infty} \mathfrak{Q}_n(t; \lambda)P_n(x; \lambda)/n!. \tag{42}$$

The series in (41, 42) converge uniformly for all t, x, λ bounded (on deleting $\lambda = \lambda_0, \lambda_1, \cdots$).

From the way in which (41) and (42) were established it is clear that they hold *formally* in the general theory of sets. Indeed we can state the more general

THEOREM 5. *Let P be any polynomial set, and let $\mathfrak{Q}$ be the associate of the inverse of P. Then*

$$e^{tx} \sim \sum_{n=0}^{\infty} P_n(x)\mathfrak{Q}_n(t)/n!. \tag{43}$$

For, let L be the differential operator* (in general of infinite order) that carries the identity set I into P: $L[x^n] = P_n(x)$, and let Q be the inverse of P, so that $L[Q_n(x)] = x^n$. Since Q and $\mathfrak{Q}$ are associate,

$$Q(t, x) \sim \sum_0^\infty Q_n(x)t^n/n! \sim \sum_0^\infty \mathfrak{Q}_n(t)x^n/n!.$$

Then,

$$L[Q(t, x)] \sim e^{tx} \sim \sum_n \mathfrak{Q}_n(t)L[x^n]/n! \sim \sum_n \mathfrak{Q}_n(t)P_n(x)/n!,$$

and this is (43).

If P is a polynomial set (if $\mathfrak{Q}$ is a triangular set) then $\mathfrak{Q}$ (P) is uniquely determined by (43). Hence we have the converse

THEOREM 6. *If P, $\mathfrak{Q}$ are any polynomial and triangular sets, respectively, satisfying* (43), *then $\mathfrak{Q}(P)$ is the associate of the inverse of $P(\mathfrak{Q})$.*

We can now complete Theorem 3. Define

$$R_n(t, x; \lambda) = (\lambda - \lambda_n)Q(t, x; \lambda).$$

---

* That L exists is established in *Sets*, p. 29.

Since $Q_i(x;\lambda)$ contains at most a simple pole at $\lambda=\lambda_n$, $R_n(t,x;\lambda)$ is analytic at $\lambda=\lambda_n$. $Q(t,x;\lambda)$ has then, at $\lambda=\lambda_n$, at most a pole of first order. Now

$$\mathcal{L}[Q(t,x;\lambda)] = \mathcal{L}_\lambda[Q] + \lambda Q = e^{tx} + \lambda Q(t,x;\lambda);$$

therefore $\mathcal{L}[R_n(t,x;\lambda)] = (\lambda-\lambda_n)e^{tx} + \lambda R_n(t,x;\lambda)$, and (on letting $\lambda\to\lambda_n$) $\mathcal{L}[R_n(t,x;\lambda_n)] = \lambda_n R_n(t,x;\lambda_n)$. Now $R_n(t,x;\lambda_n)$ can be expanded about $t=0$. Hence by Theorem 2 it can differ from $\mathcal{D}_n(t)$ at most by a factor independent of t: $R_n(t,x;\lambda_n) = H_n(x)\mathcal{D}_n(t)$. Similarly, $L[R_n(t,x;\lambda_n)] = \lambda_n R_n(t,x;\lambda_n)$, and since $R_n(t,x;\lambda_n)$ is an entire function in x, we must have (Theorem 1)

(i) $$R_n(t,x;\lambda_n) = c_n y_n(x)\mathcal{D}_n(t) \quad (c_n = \text{constant}).$$

We agree from now on that in $y_n(x)$ and $\mathcal{D}_n(t)$ we shall choose the coefficients of x^n, t^n respectively to be unity:

(ii) $$y_{nn} = 1,\ d_{nn} = 1.$$

From (29) we see that the coefficient of x^n in $(\lambda-\lambda_n)Q_n(x;\lambda)$ is -1, and since

$$R_n(t,x;\lambda_n) = \sum_{i=n}^{\infty}\left\{\lim_{\lambda\to\lambda_n}(\lambda-\lambda_n)Q_n(x;\lambda)\right\}t^n/n!,$$

the coefficient of x^nt^n in $R_n(t,x;\lambda_n)$ is $-1/n!$. Hence, by (i, ii), $c_n=-1/n!$.

Now let γ_n be a closed contour in the λ-plane surrounding the point $\lambda=\lambda_n$ but containing in its interior and on its boundary no other λ_i. Then

$$\frac{1}{2\pi i}\int_{\gamma_n} Q(t,x;\lambda)d\lambda = \lim_{\lambda\to\lambda_n}(\lambda-\lambda_n)Q(t,x;\lambda) = R_n(t,x;\lambda_n).$$

THEOREM 7. *At each of the points* $\lambda=\lambda_n$, $Q(t,x;\lambda)$ *has a simple pole with residue* $-y_n(x)\mathcal{D}_n(t)/n!$, *so that*

(44) $$\frac{1}{2\pi i}\int_{\gamma_n} Q(t,x;\lambda)d\lambda = -\,y_n(x)\mathcal{D}_n(t)/n!.$$

From (44) we have

$$-\,y_n(x)\mathcal{D}_n(t)/n! = \lim_{\lambda\to\lambda_n}(\lambda-\lambda_n)Q(t,x;\lambda).$$

On using (32), then,

(a) $$-\,y_n(x)\mathcal{D}_n(t)/n! = \lim_{\lambda\to\lambda_n}(\lambda-\lambda_n)Q_n(x;\lambda)t^n/n! + \lim_{\lambda\to\lambda_n}(\lambda-\lambda_n)Q_{n+1}(x;\lambda)t^{n+1}/(n+1)! + \cdots,$$

so that on equating coefficients of t^n,

(b) $$-y_n(x)/n! = \lim_{\lambda\to\lambda_n} (\lambda - \lambda_n)Q_n(x;\lambda)/n!.$$

From (b) follows the relation

(c) $$y_n(x) = \frac{-1}{2\pi i}\int_{\gamma_n} Q_n(x;\lambda)d\lambda.$$

Using (29) in (b), we can express $y_n(x)$ in terms of the Q_i's:

(44′) $$y_n(x) = x^n - \left\{\sigma_{n-1,n}Q_{n-1}(x;\lambda_n) + \cdots + \sigma_{n-k,n}Q_{n-k}(x;\lambda_n)\right\}.$$

This relation can be reversed rather simply as follows:

Equating like powers of t in (a) gives

$$-y_n(x)d_{n,n+i}/n! = \lim_{\lambda\to\lambda_n}(\lambda-\lambda_n)Q_{n+i}(x;\lambda)/(n+i)!,$$

so that $y_n(x)$ is, with a certain constant factor, the residue of $Q_{n+i}(x;\lambda)$ at $\lambda=\lambda_n$. Hence $y_{n-i}(x)$ is, to within a factor, the residue of $Q_n(x;\lambda)$ at $\lambda=\lambda_{n-i}$, and there exist certain constants $q_{n0}, q_{n1}, \cdots, q_{nn}$ (independent of x and λ) such that

(44″) $$Q_n(x;\lambda) = q_{n0}\frac{y_0(x)}{\lambda-\lambda_0} + q_{n1}\frac{y_1(x)}{\lambda-\lambda_1} + \cdots + q_{nn}\frac{y_n(x)}{\lambda-\lambda_n}.$$

The q_{ni}'s can be found successively from recurrence relations.

The relation (44) suggests the partial fraction expansion

(45) $$Q(t,x;\lambda) = \sum_{n=0}^{\infty}\frac{y_n(x)\mathcal{D}_n(t)}{(\lambda_n-\lambda)n!}.$$

Assuming (45) to converge uniformly, and applying the operator L_λ, we obtain

(46) $$e^{tx} = \sum_{n=0}^{\infty} y_n(x)\mathcal{D}_n(t)/n!.$$

We shall establish the validity of (45, 46) after we have developed some necessary inequalities. Let us observe, however, that (46) is of type (43) so that by Theorem 6 (and this holds for the general theory of sets) we have the

COROLLARY. *The sets* $\{y_n(x)\}$, $\{\mathcal{D}_n(t)\}$ *are respectively the associates of each other's inverse.*

4. Some inequalities for $y_n(x)$, $\mathcal{D}_n(t)$. Let λ_n be defined as in (3).

LEMMA 5. *For all n and for all $i>0$,*

$$(47) \qquad |\lambda_{n+i}-\lambda_n| \geqq C\left[\frac{k}{1!}n^{k-1}i+\frac{k(k-1)}{2!}n^{k-2}i^2+\cdots+\frac{k!}{k!}i^k\right],$$

where C is a positive constant independent of both n and i.

To show this, we can write

$$(a) \qquad \lambda_{n+i}-\lambda_n = B(n,i)+l_{kk}[A_{k-1,1}n^{k-1}i+A_{k-2,2}n^{k-2}i^2+\cdots+A_{0k}i^k]$$

where $B(n,i)$ is a polynomial in n and i of degree $<k$; and a simple calculation gives for the A_{rs} the values

$$A_{k-1,1}=k/1!,\ A_{k-2,2}=k(k-1)/2!,\cdots,A_{0k}=k!/k!.$$

The bracket on the right hand side of (a) agrees, then, with the bracket in (47). It is now a straightforward argument to show: *first*, that for n or i (or both) sufficiently large, the bracket is the dominant term in (a); *secondly*, that for i and n both bounded, a C exists. (47) then follows.

With a properly adjusted C we have the

COROLLARY. *For all n and all $i>0$,*

$$(48) \qquad |\lambda_{n+i}-\lambda_n| \geqq Ci(n+i)^{k-1} \quad (C>0, \textit{ independent of } n \textit{ and } i).$$

The coefficients in the series

$$\mathcal{D}_n(t) = \sum_{i=n}^{\infty} d_{ni}t^i$$

are given (see (13)) by*

$$(49) \qquad \begin{aligned}(\lambda_s-\lambda_n)d_s+\alpha_{s,s-1}d_{s-1}+\alpha_{s,s-2}d_{s-2}+\cdots+\alpha_{s,s-k}d_{s-k}&=0,\\ s=n+1,n+2,\cdots,&\end{aligned}$$

with $d_n=1$. Let S be a positive number such that

$$(b) \qquad |l_{ij}| \leqq S, \quad 0 \leqq i, \quad j \leqq k.$$

Then

LEMMA 6. *For all n and $i(i>0)$ we have, with C as in* (48),

$$(50) \qquad |d_{n,n+i}| \leqq \frac{1}{i!}\left(\frac{Sk}{C}\right)\left(1+\frac{Sk}{C}\right)^{i-1}.$$

(50) is readily established (by use of (48)) for $i=1$, 2. Assuming it to be true up to $i-1$, we shall prove it true for i by induction, as follows:

* For simplicity we use d_s for d_{ns}.

$$| d_{n+i} | \leqq \frac{1}{| \lambda_{n+i} - \lambda_n |}[| \alpha_{n+i,n+i-1}d_{n+i-1} | + \cdots + | \alpha_{n+i,n+i-k}d_{n+i-k} |]$$

$$\leqq \frac{Sk}{Ci(n+i)^{k-1}}\left[(n+i-1)^{k-1}\frac{db^{i-2}}{(i-1)!} + (n+i-2)^{k-2}\frac{db^{i-3}}{(i-2)!}\right.$$

$$\left. + \cdots + (n+i-k)^0\frac{db^{i-k-1}}{(i-k)!}\right],$$

where $a = (Sk/C)$, $b = 1 + (Sk/C)$;

$$| d_{n+i} | \leqq \frac{a^2b^{i-k-1}}{i!}\left[b^{k-1} + \frac{(i-1)}{n+i}b^{k-2} + \frac{(i-1)(i-2)}{(n+i)^2}b^{k-3} + \cdots\right.$$

$$\left. + \frac{(i-1)\cdots(i-k+1)}{(n+i)^{k-1}}\right]$$

$$\leqq \frac{a^2b^{i-k-1}}{i!}[b^{k-1} + b^{k-2} + \cdots + 1] \leqq \frac{b^k}{b-1}\cdot\frac{a^2b^{i-k-1}}{i!} = \frac{ab^{i-1}}{i!},$$

and this is (50).

Consequently,

Lemma 7.

$$\mathcal{D}_n(t) \ll t^n + \frac{a}{1!}t^{n+1} + \frac{ab}{2!}t^{n+2} + \cdots + \frac{ab^{i-1}}{i!}t^{n+i} + \cdots \ll pt^ne^{bt}, \tag{51}$$

a, b, $p = \max(1, a)$ independent of n.

We now turn to $y_n(x)$. Its coefficients satisfy (as we see from (5)) the equations

$$(\lambda_s - \lambda_n)y_s + \sigma_{s,s+1}y_{s+1} + \sigma_{s,s+2}y_{s+2} + \cdots + \sigma_{s,s+k}y_{s+k} = 0 \qquad (s = 0, 1, \cdots, n-1), \tag{52}$$

with $y_n = 1$.

Lemma 8. *The coefficients of $y_n(x)$ satisfy the inequality*

$$| y_{n,n-i} | \leqq \frac{n(n-1)\cdots(n-i+1)}{i!}\left(\frac{Sk}{C}\right)\left(1 + \frac{Sk}{C}\right)^{i-1} \tag{53}$$

for all n and i ($i \leqq n$).

The proof, by induction, differs very little from that of Lemma 6, and may be omitted. From (53) we get

LEMMA 9.

$$
(54) \quad \begin{aligned} y_n(x) &\ll x^n + \frac{n}{1!}ax^{n-1} + \frac{n(n-1)}{2!}abx^{n-2} + \cdots + \frac{n!}{n!}ab^{n-1} \\ &\ll q\left[x^n + \frac{n}{1!}x^{n-1}b + \cdots + \frac{n!}{n!}b^n\right] \ll q(x+b)^n, \end{aligned}
$$

$q = \max(1, a/b)$ *independent of* n, *so that for all* n *and* x,

$$
(55) \quad |y_n(x)| \leqq q(|x| + b)^n.
$$

Combining Lemmas 7, 9:

$$
|y_n(x)\mathcal{D}_n(t)| \leqq pqe^{b|t|}[|t|(|x| + b)]^n \text{ (for all } n, x, t).
$$

Hence we have

THEOREM 8. *The series* $\sum_0^\infty y_n(x)\mathcal{D}_n(t)/n!$ *converges uniformly in every bounded* x, t *region, and represents an entire function in the two variables* x, t.

COROLLARY. *The series* $\sum_0^\infty \{y_n(x)\mathcal{D}_n(t)/[(\lambda_n - \lambda)n!]\}$ *converges uniformly in every bounded* x, t, λ *region (the points* $\lambda = \lambda_0, \lambda_1, \cdots$ *being deleted).*

The above two series are the right hand members of (45, 46). It remains to prove that they represent the corresponding left hand members. Let $H(t, x; \lambda)$ denote the sum of the series in the above corollary. H and Q have then the same principal parts at $\lambda = \lambda_0, \lambda_1, \cdots$ so that $Q - H$ is an entire function in all three variables. On applying the operator L_λ term-wise (as we may) to the series H, we get

$$
\text{(a)} \quad L_\lambda[Q(t, x; \lambda) - H(t, x; \lambda)] = e^{tx} - \sum_0^\infty y_n(x)\mathcal{D}_n(t)/n!.
$$

Hence the left hand member is independent of λ, and represents a function $C(t, x)$ that is entire in t and x: $C(t, x) = \sum_{m,n=0}^\infty c_{mn}x^m t^n$. The right hand member of (a) has *zero* as coefficient of $x^m t^n$, $m \geqq n$, so that $c_{mn} = 0$, $m \geqq n$. Hence

$$
\text{(b)} \quad C(t, x) = \sum_{n=1}^\infty c_n(x)t^n,
$$

where $c_n(x)$ is a polynomial of degree not exceeding $n - 1$.

Now e^{tx} and $\sum_0^\infty y_n(x)\mathcal{D}_n(t)/n!$ are self-dual functions; the same is then true of $C(t, x)$:

$$
\text{(c)} \quad L[C(t, x)] = \mathcal{L}[C(t, x)].
$$

On substituting into (c) the series (b), and equating coefficients of like powers of t, we obtain the equations

$$\text{(d)}\qquad \begin{aligned} &L_0(x)c_n(x) + L_1(x)c_n'(x) + \cdots + L_k(x)c_n^{(k)}(x) \\ &= \lambda_n c_n(x) + \alpha_{n,n-1}c_{n-1}(x) + \alpha_{n,n-2}c_{n-2}(x) + \cdots + \alpha_{n,n-k}c_{n-k}(x) \\ &\qquad\qquad (n = 1, 2, \cdots). \end{aligned}$$

It is at once verified that† $c_1(x)=0$. Assume that $c_1(x)=c_2(x)=\cdots=c_{n-1}(x)=0$. We shall show that $c_n(x)=0$. On our induction assumption, (d) reduces to

$$\text{(e)}\qquad L[c_n(x)] = \lambda_n c_n(x).$$

$c_n(x)$ is an entire function, so that by Theorem 1, $c_n(x)\equiv a_n y_n(x)$, a_n a constant. But $c_n(x)$ is of degree less than n; hence $a_n=0$, and $c_n(x)=0$. That is, $C(t, x)\equiv 0$, and the right hand member of (a) is zero.

We have just established (46). Equation (a) then gives us

$$\text{(a}'\text{)}\qquad L[Q(t, x; \lambda) - H(t, x; \lambda)] = 0,$$

where $Q-H$ is entire in t, x, λ. By Theorem 1, (a′) has an entire function solution ($\not\equiv 0$) if and only if $\lambda=\lambda_0, \lambda_1, \cdots$. Hence $Q-H\equiv 0$, $\lambda\neq\lambda_0, \lambda_1, \cdots$, and by continuity $Q-H\equiv 0$ for all t, x, λ. That is, $Q\equiv H$, and this is (45). We thus have

THEOREM 9. *The two series*

$$\text{(45)}\quad Q(t, x; \lambda) = \sum_{n=0}^{\infty} \frac{y_n(x)\mathcal{D}_n(t)}{(\lambda_n - \lambda)n!}, \qquad \text{(46)}\quad e^{tx} = \sum_{n=0}^{\infty} y_n(x)\mathcal{D}_n(t)/n!$$

are valid for all t, x, λ ($\lambda_0, \lambda_1, \cdots$ deleted).

COROLLARY. *The expansion*

$$\text{(56)}\qquad P(t, x; \lambda) = \sum_{n=0}^{\infty} (\lambda_n - \lambda) y_n(x)\mathcal{D}_n(t)/n!$$

is uniformly convergent in every bounded t, x, λ region, thus representing an entire function in all three variables. (See (26).)

5. **Further inequalities; expansions in $\mathcal{D}_n(t)$, $Y_n(x)$.** In the present section we investigate the question of expansions of functions in terms of the two sets $\{\mathcal{D}_n(t)\}$, $\{Y_n(x)\}$ (introduced in (19)). For this we require some further inequalities. In the equation

† For $c_1(x)\equiv$ constant $=c$, say. Then $l_{00}c=(l_{00}+l_{11})c$, so that $cl_{11}=0$. Now if $l_{11}=0$, then $\lambda_1=\lambda_0$, which contradicts our assumption ($\lambda_m\neq\lambda_n$, $m\neq n$). Hence $l_{11}\neq 0$, and $c=0$.

$$(4') \qquad L[y_n(x)] = \lambda_n y_n(x)$$

make the transformation $x = x^* - \gamma$. On setting $y_n^*(x^*) = y_n(x)$ we have

$$(4'') \qquad L^*[y_n^*(x^*)] = \lambda_n y_n^*(x^*),$$

where L^* is an operator similar to L, $L_i^*(x^*)$ being equal to $L_i(x)$. In particular, $l_{kk}^* = l_{kk}$, $l_{k,k-1}^* = -kl_{kk}\gamma + l_{k,k-1}$.

If we denote by $\{\mathcal{D}_n^*(t)\}$ the set of functions corresponding to $\{y_n^*(x^*)\}$, then by (46),

$$(46') \qquad e^{tx^*} = \sum_{n=0}^{\infty} y_n^*(x^*)\mathcal{D}_n^*(t)/n!.$$

But $y_n^*(x^*) = y_n(x)$, $e^{tx^*} = e^{tx} \cdot e^{\gamma t}$. Hence $\mathcal{D}_n(t)$ is transformed into

$$(57) \qquad \mathcal{D}_n^*(t) \equiv e^{\gamma t}\mathcal{D}_n(t).$$

Choose

$$(58) \qquad \gamma = l_{k,k-1}/kl_{kk}.$$

Then $l_{k,k-1}^* = 0$; i.e., the sum of all the zeros of $L_k^*(x^*)$ is zero. There is clearly no loss in generality if we go from (4) to (4″) for the choice of γ in (58). Then, dropping asterisks, *we shall henceforth suppose that in* (4′), $l_{k,k-1} = 0$.

LEMMA 10. *For all n and $s(s>0)$ we have*

$$(59) \qquad |d_{n,n+s}| \leqq \frac{1}{n}\frac{h^s}{s!}, \quad d_{nn} = 1,$$

h being independent of n and s.

To show this, substitute in equations (49) for the d's the numbers r_i defined by

$$(a) \qquad r_{n+s} = nd_{n+s},\ s = 0, 1, \cdots, r_n = n.$$

This gives

$$(b) \qquad (\lambda_{n+s} - \lambda_n)r_{n+s} + \alpha_{n+s,n+s-1}r_{n+s-1} + \cdots + \alpha_{n+s,n+s-k}r_{n+s-k} = 0 \qquad (s = 1, 2, \cdots).$$

Solving for r_{n+s}, and using the values of the α_{ij}'s (given by (14)) as well as the relation $l_{k,k-1} = 0$ and the inequalities (48), we find that

$$(c) \qquad r_n = n, \ |r_{n+s}| \leqq (h/s)\left[\frac{|r_{n+s-1}|}{n+s} + \frac{|r_{n+s-2}|}{n+s} + \frac{|r_{n+s-3}|}{(n+s)^2} + \cdots + \frac{|r_{n+s-k}|}{(n+s)^{k-1}}\right],$$

where $h \geqq Sk/C$. Choose $h = \max(2, Sk/C)$. On setting

$$\text{(d)} \quad t_n = n,\ t_{n+s} = (h/s)\left[\frac{t_{n+s-1}}{n+s} + \frac{t_{n+s-2}}{(n+s)} + \frac{t_{n+s-3}}{(n+s)^2} + \cdots + \frac{t_{n+s-k}}{(n+s)^{k-1}}\right],$$

we have $t_{n+s} \geqq |r_{n+s}|$.

A simple calculation gives us $t_{n+1} \leqq h/1!$, $t_{n+2} \leqq h^2/2!$; we shall establish the relation

$$\text{(e)} \qquad t_{n+s} \leqq h^s/s!, \quad s > 0,$$

by induction, assuming it true up to $s-1$. For

$$\begin{aligned} t_{n+s} &\leqq \frac{h}{s}\left[\frac{h^{s-1}}{(s-1)!(n+s)} + \frac{h^{s-2}}{(s-2)!(n+s)} + \cdots + \frac{h^{s-k}}{(s-k)!(n+s)^{k-1}}\right] \\ &\leqq \frac{h^s}{s!}\,\frac{1}{n+s}\left[1 + \frac{s-1}{h}\left\{1 + \frac{s-2}{h(n+s)} + \frac{(s-2)(s-3)}{h^2(n+s)^2}\right.\right. \\ &\qquad \left.\left. + \cdots + \frac{(s-2)\cdots(s-k+1)}{h^{k-2}(n+s)^{k-2}}\right\}\right] \\ &\leqq \frac{h^s}{s!}\,\frac{1}{n+s}\left[1 + \frac{2(s-1)}{h}\right] \leqq \frac{h^s}{s!}, \end{aligned}$$

which is (e). Then, $|r_{n+s}| \leqq h^s/s!$, and from this (59) follows. From (59) we get

THEOREM 10. *$\mathcal{D}_n(t)$ is asymptotically given by*

$$\text{(60)} \qquad \mathcal{D}_n(t) = t^n[1 + A_n(t)]$$

where

$$\text{(61)} \qquad |A_n(t)| \leqq e^{h|t|}/n.$$

THEOREM 11. *Let $C(t) = \sum_0^\infty c_n t^n$ have r as its radius of convergence. Then the series*† *$C^*(t) = \sum_0^\infty c_n \mathcal{D}_n(t)$*

(a) *converges absolutely at every point in $|t| < r$;*

(b) *converges uniformly in $|t| \leqq r' < r$, r' arbitrary;*

(c) *diverges in $|t| > r$.*

In particular, $\mathcal{D}_n(t)$-expansions have circles, center at origin, as their regions of convergence.

In fact, for n sufficiently large,

$$\tfrac{1}{2}|c_n t^n| \leqq |c_n \mathcal{D}_n(t)| = |c_n t^n| \cdot |1 + A_n(t)| \leqq 2|c_n t^n|.$$

† If $r = 0$, the only point of convergence for the $C^*(t)$-series is $t = 0$.

In $C^*(t)=\sum_0^\infty c_n\mathcal{D}_n(t)$, let $\limsup |c_n|^{1/n}=\sigma<\infty$, so that the series has a radius of convergence $r=1/\sigma>0$. We may expand $C^*(t)$ in a power series in t: $C^*(t)=\sum_0^\infty c_n^* t^n$, with

$$\text{(i)} \qquad c_n^* = c_0 d_{0n} + c_1 d_{1n} + \cdots + c_n d_{nn} \qquad (n = 0, 1, \cdots).$$

Since, by hypothesis, $|c_n| \leqq A(\sigma+\epsilon)^n$, $\epsilon>0$, $A=A_\epsilon$, we have (see (50))

$$|c_n^*| \leqq A[a(1+a)^{n-1}/n! + (\sigma+\epsilon)a(1+a)^{n-2}/(n-1)! + \cdots + (\sigma+\epsilon)^{n-1}a/1! + (\sigma+\epsilon)^n],$$

$a=Sk/C$. That is,

$$|c_n^*| \leqq A(\sigma+\epsilon)^n\left[1+\frac{a}{\sigma+\epsilon}\left\{\frac{1}{1!}+\frac{1}{2!}\left(\frac{1+a}{\sigma+\epsilon}\right)+\cdots+\frac{1}{n!}\left(\frac{1+a}{\sigma+\epsilon}\right)^{n-1}\right\}\right]$$

$$\leqq A\left[1+\frac{a}{\sigma+\epsilon}e^{(1+a)/(\sigma+\epsilon)}\right](\sigma+\epsilon)^n \leqq B(\sigma+\epsilon)^n, \ B \text{ independent of } n.$$

Therefore $\limsup |c_n^*|^{1/n}\leqq\sigma$, and we have

LEMMA 11. *If $C^*(t)=\sum_0^\infty c_n\mathcal{D}_n(t)$ has the radius of convergence r, then $C^*(t)=\sum_0^\infty c_n^*t^n$ has a radius of convergence at least as great as r.*

The converse theorem is also true. Its proof, which is not so immediate, can be made to depend on inequalities regarding the functions $Y_n(x)$ of (19):

$$\text{(19)} \qquad Y_n(x) = 0!y_{n0} + 1!y_{n1}x + \cdots + n!y_{nn}x^n.$$

Set $z_{ni}=i!y_{ni}/n!$. Then, by (5),

$$\text{(62)} \qquad \begin{aligned}(\lambda_s-\lambda_n)(s+k)\cdots(s+1)z_s + \sigma_{s,s+1}(s+k)\cdots(s+2)z_{s+1} + \cdots \\ + \sigma_{s,s+k}z_{s+k} = 0 \quad (s = 0, 1, \cdots),\end{aligned}$$

with $z_n=1$. We find, using (48), that

$$|z_{n-1}| \leqq \frac{h}{1!n}, \quad |z_{n-2}| \leqq \frac{h^2}{2!n}, \quad n>1, \ h=\max(2, Sk/C),$$

and an induction gives us (compare Lemma 10)

LEMMA 12. *For all $n>1$ and all i $(0<i\leqq n)$,*

$$\text{(ii)} \qquad |z_{n,n-i}| \leqq h^i/(i!n).$$

We can write

$$Y_n(x)/n! = x^n[z_n + z_{n-1}/x + \cdots + z_0/x^n]$$

$$= x^n\left[1+\frac{1}{n}\left\{\frac{1}{1!}\left(\frac{h}{x}\right)\delta_{n1}+\frac{1}{2!}\left(\frac{h}{x}\right)^2\delta_{n2}+\cdots+\frac{1}{n!}\left(\frac{h}{x}\right)^n\delta_{nn}\right\}\right],$$

where $|\delta_{ni}| \leqq 1$ for all n and i. This gives us

THEOREM 12. *$Y_n(x)$ has the asymptotic form*

$$Y_n(x)/n! = x^n[1 + B_n(x)/n] \tag{63}$$

where†

$$|B_n(x)| \leqq e^{h/|x|}. \tag{64}$$

Let us now consider the expansion (46) for e^{tx}. This is (in the variable x) the Borel entire function associated with $1/(1-tx)$, and $y_n(x)$ is the Borel function associated with $Y_n(x)$, thus suggesting the expansions

$$1/(1 - tx) = \sum_{n=0}^{\infty} Y_n(x)\mathcal{D}_n(t)/n! \equiv T(t, x), \tag{a}$$

$$\frac{1}{x - t} = \sum_{n=0}^{\infty} \frac{Y_n(1/x)\mathcal{D}_n(t)}{x(n!)}. \tag{65}$$

To verify this and determine the region of validity of (a, 65), we appeal to relations (60) and (63), which show us that $T(t, x)$ is analytic in any region for which $|tx|<1$, and that the series for $T(t, x)$ converges uniformly for $|tx| \leqq \alpha < 1$. Let x trace the circle $|x| = \delta > 0$. The series in question then converges uniformly for $|t| \leqq \alpha/\delta$, and we may multiply it by $e^{x/u}/u$ and integrate term-wise with respect to u around Γ: $|u| = \delta$:

$$\frac{1}{2\pi i}\int_\Gamma \frac{T(t, u)}{u} e^{x/u} du = \sum_0^\infty \frac{\mathcal{D}_n(t)}{n!}\left\{\frac{1}{2\pi i}\int_\Gamma \frac{Y_n(u)}{u} e^{x/u} du\right\}, \quad = e^{tx}$$

since

$$\frac{1}{2\pi i}\int_\Gamma \frac{Y_n(u)}{u} e^{x/u} du = y_n(x).$$

If we now expand $T(t, u)$ in a power series about $u=0$ (as we may):

$$T(t, u) = \sum_0^\infty T_n(t)u^n,$$

we find that $T_n(t) = t^n$, so that (a) is true for $|tx|<1$. Therefore (65) holds. Now by (63), $\lim\sup|Y_n(1/x)/n!|^{1/n} = 1/|x|$; whence from Theorem 11 follows

† For $x=0$, $|Y_n(x)/n!| = |z_{n0}| \leqq h^n/(n!\,n)$.

THEOREM 13. *For every x the series* (65) *has the interior of the circle* $|t| = |x|$ *as its region*† *of convergence. In every region* $|t| \leqq \rho < |x|$ *the convergence is uniform.*

If $f(t) = \sum_0^\infty f_n t^n$ is analytic in $|t| < r$, we get from (65), by Cauchy's integral formula:

$$f(t) = \sum_{n=0}^{\infty} \left\{ \frac{1}{2\pi i} \int_{|x|=r'} \frac{Y_n(1/x) f(x)}{x(n!)} dx \right\} \mathcal{D}_n(t), \tag{66}$$

valid for $|t| < r' < r$. But r' can be chosen as close to r as we desire; hence (66) is true for all $|t| < r$, and is uniformly convergent in every closed region in $|t| < r$. This is the converse of Lemma 11. Combining the two we have

THEOREM 14. *The function* $f(t)$ *has a convergent* $\mathcal{D}_n(t)$*-expansion if and only if it is analytic about* $t = 0$; *and its* $\mathcal{D}_n(t)$*-expansion and its power series expansion have the same radius of convergence.*

A $\mathcal{D}_n(t)$*-expansion is unique.* For if $\sum_0^\infty c_n \mathcal{D}_n(t)$ converges, it converges uniformly (Theorem 11), and we may write $\sum_0^\infty c_n \mathcal{D}_n(t) = \sum_0^\infty c_n^* t^n$, where the c_n^* are given by (i). If the c_n^*'s are given, these equations (i) determine $c_0, c_1, \cdots$ uniquely.

Let us sum up our theorems on $\mathcal{D}_n(t)$-series:

THEOREM 15. *The series* $\sum_0^\infty c_n \mathcal{D}_n(t)$ *has the single point of convergence* $t = 0$ *if and only if* $\limsup |c_n|^{1/n} = \infty$. *If* $\limsup |c_n|^{1/n} = \sigma < \infty$, *then the series converges throughout* $|t| < 1/\sigma$ *and diverges throughout* $|t| > 1/\sigma$. *In* $|t| < 1/\sigma$ *the convergence is absolute, and in every closed region in* $|t| < 1/\sigma$ *it is uniform. If* $f(t)$ *denotes the sum of the series, then* $f(t)$ *is analytic in* $|t| < 1/\sigma$, *and* $1/\sigma$ *is the radius of convergence of* $f(t) = \sum_0^\infty f_n t^n$, *i.e., the* $\mathcal{D}_n(t)$*-series and the power series for the same function have the same radius of convergence. Furthermore, a* $\mathcal{D}_n(t)$*-expansion is unique, and the coefficients are given by*

$$c_n = \frac{1}{2\pi i} \int_C \frac{f(x)}{x} \cdot \frac{Y_n(1/x)}{n!} dx, \tag{67}$$

C being a contour about $x = 0$ *and lying in* $|x| < 1/\sigma$.

We turn now to $Y_n(x)$-expansions. From (60, 63, 46) we readily deduce

LEMMA 13. *The function* $1/(t - x)$ *has the expansion*

$$\frac{1}{t - x} = \sum_{n=0}^{\infty} \frac{Y_n(x) \mathcal{D}_n(1/t)}{t(n!)}, \tag{68}$$

† Points of the boundary may be points of convergence.

which for a given t has the interior of the circle $|x| = |t|$ *as its region of convergence. In every region* $|x| \leqq \rho < |t|$ *the convergence is uniform.*

In ways analogous to those used for Theorem 15, we can establish

THEOREM 16. *Everything said*† *of* $\mathcal{D}_n(t)$*-expansions in Theorem* 15 *holds for* $\{Y_n(x)/n!\}$*-expansions, with the modification that the* c_n*'s in* $\sum_0^\infty c_n Y_n(x)/n!$ *are now given by*

$$c_n = \frac{1}{2\pi i}\int_C \frac{f(t)}{t}\,\mathcal{D}_n(1/t)dt, \tag{69}$$

C being a contour about $t=0$ *and lying in* $|t| < 1/\sigma$.

6. **Biorthogonality relations; differential equations for** $\Delta_n(t)$, $Y_n(x)$. In the equation

$$\mathcal{D}_n(t) = \frac{1}{2\pi i}\int_C \frac{e^{ut}}{u}\,\Delta_n(1/u)du, \tag{18}$$

C being a contour surrounding $u=0$ and lying outside of $|u| = \rho$, replace e^{ut} by its expansion $\sum_{s=0}^\infty y_s(u)\mathcal{D}_s(t)/s!$, which converges uniformly on C. This gives us

$$\mathcal{D}_n(t) = \sum_{s=0}^{\infty}\left\{\frac{1}{2\pi i}\int_C \frac{y_s(u)\Delta_n(1/u)}{s!u}du\right\}\mathcal{D}_s(t),$$

whence by uniqueness of $\mathcal{D}_n(t)$-expansions we have

THEOREM 17. *The functions* $\{y_n\}$, $\{\Delta_n\}$ *are biorthogonal in the following sense*:

$$\frac{1}{2\pi i}\int_C \frac{y_s(u)\Delta_n(1/u)}{n!u}du = \begin{cases} 0, & s \neq n, \\ 1, & s = n. \end{cases} \tag{70}$$

If we start with the relation

$$y_n(x) = \frac{1}{2\pi i}\int_\Gamma \frac{e^{xu}}{u}\,Y_n(1/u)du, \tag{20}$$

Γ being a contour around $u=0$, we obtain the uniformly convergent expansion

$$y_n(x) = \sum_{s=0}^{\infty}\left\{\frac{1}{2\pi i}\int_\Gamma \frac{\mathcal{D}_s(u)Y_n(1/u)}{s!u}du\right\}y_s(x). \tag{a}$$

As we have not established uniqueness of $y_n(x)$-expansions, we cannot at once conclude that the brace in (a) is zero or one. But this can be proved in the following way.

† We must of course except the conclusion of Theorem 15 for the case $\limsup_{n\to\infty} |c_n|^{1/n} = \infty$.

Denote by c_{ns} the brace in (a), so that $y_n(u)=\sum_{s=0}^{\infty}c_{ns}y_s(u)$, uniformly convergent for all bounded u. Multiply both members by $\Delta_r(1/u)/(n!u)$ and integrate term-wise over the contour C of (70). This gives us, by (70),

$$\delta_{nr} = \sum_{s=0}^{\infty} c_{ns}s!\delta_{sr}/n!; \quad \text{i.e.}, \delta_{nr} = \frac{c_{nr}r!}{n!},$$

or $c_{nr}=0$, $r\neq n$; $c_{nn}=1$. Hence we have

THEOREM 18. *The functions* $\{\mathcal{D}_n\}$, $\{Y_n\}$ *are biorthogonal in the sense*

$$\frac{1}{2\pi i}\int_{\Gamma}\frac{\mathcal{D}_s(u)Y_n(1/u)}{n!u}du = \begin{cases} 0, & s\neq n, \\ 1, & s=n. \end{cases} \tag{71}$$

By means of (70) we can show that a $y_n(x)$-*expansion is unique if it converges uniformly in a region that contains the region* $|x|\leqq\rho+\epsilon,\epsilon>0$ sufficiently small. Here ρ is, as before, the maximum absolute value of the zeros of $L_k(x)$. This is equivalent to saying that if the function *zero* has a y_n-expansion that is uniformly convergent in $|x|\leqq\rho+\epsilon$, then the coefficients in the expansion are all zero. This result follows on multiplying the series in question through by $(1/u)(1/n!)\Delta_n(1/u)$ and integrating over C, using (70).

We can sharpen this conclusion by further considering the functions $\Delta_n(t)$. We know $\mathcal{D}_n(t)$ is the Borel entire function corresponding to $\Delta_n(t)$. In general,

DEFINITION. *If the two functions* $A(t)=\sum_0^{\infty}a_nt^n$, $B(t)=\sum_0^{\infty}n!a_nt^n$ *are analytic about* $t=0$, *then* $A(t)$ *is the Borel entire function associated with* $B(t)$, *and we write* $A(t)=BEF\{B(t)\}$.

It is easily established that

LEMMA 14. *If* $A(t)=BEF\{B(t)\}$, *then* $A'(t)=BEF\{(B(t)-B(0))/t\}$.

LEMMA 15. *For all* $0\leqq j\leqq i$,

$$t^i\mathcal{D}_n^{(j)}(t) = BEF\left\{t^i\frac{d^i}{dt^i}(t^{i-j}\Delta_n(t))\right\}. \tag{72}$$

This can be established by a straightforward induction argument, using Lemma 14; the proof may therefore be omitted.

Since $\mathcal{D}_n(t)$ satisfies the differential equation

$$\mathcal{L}[\mathcal{D}_n(t)] \equiv \mathcal{L}_0(t)\mathcal{D}_n(t) + \mathcal{L}_1(t)\mathcal{D}_n'(t) + \cdots + \mathcal{L}_k(t)\mathcal{D}_n^{(k)}(t) = \lambda_n\mathcal{D}_n(t),$$

where $\mathcal{L}_i(t)$ is a polynomial having no term t^k with $k<i$, we may apply Lemma 15. It gives us

THEOREM 19. *The functions $\Delta_n(t)$ satisfy the kth-order linear homogeneous differential equation*

$$\sum_{i=j}^{k}\sum_{j=0}^{k}\left\{l_{ij}t^i\frac{d^i}{dt^i}(t^{i-j}\Delta_n(t))\right\} = \lambda_n\Delta_n(t). \tag{73}$$

Here the coefficients of the various derivatives of $\Delta_n(t)$ are polynomials in t, that of $\Delta_n^{(k)}(t)$ being

$$t^k(l_{k0}t^k + l_{k1}t^{k-1} + \cdots + l_{kk}) = t^{2k}L_k(1/t).$$

COROLLARY 1. *The only possible singularities (in the finite plane) of the functions $\Delta_n(t)$ are at the reciprocals of the zeros of $L_k(t)$.*

$\Delta_n(1/t)$ has then only the zeros of $L_k(t)$, and the origin, as possible singularities. From this follows

COROLLARY 2. *In relations* (18) *and* (70), *the contour C may be chosen as any contour which has the origin and all the zeros of $L_k(u)$ in its interior.*

By the argument used just before Lemma 14, applied to equation (70) with a contour C of Corollary 2, we have

THEOREM 20. *A $y_n(x)$-expansion is unique*† *if it converges uniformly in a simply-connected open region $\mathfrak{R}$ that contains the origin and all the zeros of $L_k(x)$.*

The numbers $\{\lambda_n\}$ are the characteristic numbers for our original equations (1), (12). It is natural to inquire if they have like significance for equation (73), regarded independently of its origin. The answer, in the affirmative, is given by

THEOREM 21. *The differential equation*

$$\sum_{i=j}^{k}\sum_{j=0}^{k}\left\{l_{ij}t^i\frac{d^i}{dt^i}(t^{i-j}\Delta(t))\right\} = \lambda\Delta(t) \tag{73'}$$

has a formal power series solution about the origin if and only if $\lambda=\lambda_0, \lambda_1, \cdots$, and when $\lambda=\lambda_n$, there is a unique solution (to within an arbitrary constant multiplier); this solution converges about $t=0$, and is, in fact, the function $\Delta_n(t)$.

To prove this, assume the expansion $\Delta(t)=\sum_0^\infty a_nt^n$. On substituting into (73′) and equating coefficients, the values $\lambda_0, \lambda_1, \cdots$ are found to be the only possible ones, and these yield unique solutions. The remainder of the theorem follows from the fact that $\Delta_n(t)$ *is* a solution for $\lambda=\lambda_n$.

† That is, if two such expansions (uniformly convergent in $\mathfrak{R}$) represent the same function, corresponding coefficients are equal.

We now turn to the functions $\{Y_n(x)\}$. From equation (4) for $y_n(x)$ and the property that $y_n(x)=BEF\{Y_n(x)\}$, we derive the corresponding differential equation for $Y_n(x)$. Unlike the case for $\Delta_n(t)$, however, the new equation is of infinite order. In fact, we have

$$\text{(a)} \qquad Y_n(x) = Y_0 + Y_1 x + \cdots + Y_n x^n,$$

where $Y_i = i! y_i$. (We write Y_i, y_i for Y_{ni}, y_{ni}.) From (a) follows

LEMMA 16. *The coefficients of $Y_n(x)$ are given by*†

$$\text{(b)} \quad Y_{n-i} = [1/(n-i)!][Y_n^{(n-i)}(x) - (x/1!)Y_n^{(n-i+1)}(x) + (x^2/2!)Y_n^{(n-i+2)}(x) - \cdots + (-1)^i(x^i/i!)Y_n^{(n)}(x)] \quad (i = 0, 1, \cdots, n).$$

To show this let $T_n(x)$ denote the right hand member of (b). Since $Y_n(x)$ is of degree n, $T_n(x)$ is unaltered if we add to it terms containing higher derivatives of $Y_n(x)$ than the nth. That is, we can write

$$T_n(x) = \left[\frac{1}{(n-i)!}\right] \sum_{s=0}^{\infty}(-1)^s\left(\frac{x^s}{s!}\right)Y_n^{(n-i+s)}(x),$$

the series being uniformly convergent in every bounded region. Letting C be a contour surrounding the origin, we have

$$Y_n(x) = \frac{1}{2\pi i}\int_C \frac{Y_n(u)}{u-x}du, \quad Y_n^{(n-i+s)}(x) = \frac{(n-i+s)!}{2\pi i}\int_C \frac{Y_n(u)}{(u-x)^{n-i+s+1}}du;$$

and on substituting into $T_n(x)$, we obtain

$$T_n(x) = \frac{1}{2\pi i}\int_C \frac{Y_n(u)}{(u-x)^{n-i+1}} \cdot\left\{\sum_{s=0}^{\infty}\frac{(-1)^s}{s!}(n-i+1)(n-i+2)\cdots(n-i+s)\left(\frac{x}{u-x}\right)^s\right\}du.$$

Now the brace is the expansion of

$$\left[1+\left(\frac{x}{u-x}\right)\right]^{-(n-i+1)} = \left(\frac{u-x}{u}\right)^{n-i+1},$$

valid for $|x/(u-x)|<1$. We can choose C to satisfy this condition and also to surround the origin. Then,

$$T_n(x) = (1/(2\pi i))\int_C \{Y_n(u)/u^{n-i+1}\}du = Y_{n-i}.$$

That is, (b) holds.

† (b) is of course true for *any* polynomial of degree n.

Now let $0 \leqq j \leqq i$. From $y_n(x) = \sum_{i=0}^n y_i x_i$ we obtain

$$\begin{aligned} x^j y_n^{(i)}(x) &= n(n-1)\cdots(n-i+1)y_n x^{n-i+j} \\ &\quad + (n-1)\cdots(n-i)y_{n-1}x^{n-i+j-1} + \cdots + i!y_i x^j \\ &= BEF\left\{\frac{(n-i+j)!}{(n-i)!}Y_n x^{n-i+j} + \frac{(n-i+j-1)!}{(n-i-1)!}Y_{n-1}x^{n-i+j-1}\right. \\ &\quad \left. + \cdots + \frac{j!}{0!}Y_i x^j\right\}; \end{aligned}$$

and on using (b) of Lemma 16, this gives

$$\begin{aligned} \text{(c)} \quad x^j y_n^{(i)}(x) &= BEF\{\theta_{n;ij}x^{n-i+j}Y_n^{(n)}(x) + \theta_{n-1;ij}x^{n-i+j-1}Y_n^{(n-1)}(x) + \cdots \\ &\quad + \theta_{i;ij}x^j Y_n^{(i)}(x)\}, \end{aligned}$$

$$\begin{aligned} \text{(d)} \quad \theta_{p;ij} &= \frac{(p-i+j)!}{(p-i)!p!0!} - \frac{(p-i+j-1)!}{(p-i-1)!(p-1)!1!} + \cdots \\ &\quad + (-1)^{p-i}\frac{j!}{0!i!(p-i)!} \qquad (p = i, i+1, \cdots, n). \end{aligned}$$

We readily get

LEMMA 17. *The quantity $\theta_{p;ij}$ is the coefficient of u^{p-i} in the power series expansion of $e^{-u}H_{ij}(u)$, where $H_{ij}(u)$ is the entire function*

$$H_{ij}(u) = \sum_{s=0}^{\infty} \frac{(j+s)!}{s!(i+s)!}u^s.$$

If we substitute the value of $x^j y_n^{(i)}(x)$ as given by (c) into equation (4), for $\lambda = \lambda_n$, we obtain

$$\begin{aligned} &\sum_{i=0}^{k}\sum_{j=0}^{i} l_{ij}\{\theta_{n;ij}x^{n-i+j}Y_n^{(n)}(x) + \theta_{n-1;ij}x^{n-i+j-1}Y_n^{(n-1)}(x) + \cdots \\ \text{(74)} \quad &\qquad + \theta_{i;ij}x^j Y_n^{(i)}(x)\} = \lambda_n Y_n(x), \end{aligned}$$

which is a linear homogeneous differential equation of order n for $Y_n(x)$. (74) can be written in the form

$$\text{(74')} \quad M_0(x)Y_n(x) + M_1(x)Y_n'(x) + \cdots + M_n(x)Y_n^{(n)}(x) = \lambda_n Y_n(x),$$

where

$$\text{(75)} \quad M_i(x) = \sum_{r=0}^{s}\{l_{r0}\theta_{i;r0}x^{i-r} + l_{r1}\theta_{i;r1}x^{i-r+1} + \cdots + l_{rr}\theta_{i;rr}x^i\},$$

with $s = k$ if $i \geqq k$, and $s = i$ if $i < k$.

Clearly, $M_i(x)$ is independent of n. Since $Y_n^{(s)}(x) \equiv 0$, $s > n$, we see that *the functions* $\{Y_n(x)\}$ *are solutions of the linear homogeneous differential equation of infinite order*

$$\sum_{s=0}^{\infty} M_s(x) Y^{(s)}(x) = \lambda Y(x), \tag{76}$$

where for $Y(x) = Y_n(x)$ *we have* $\lambda = \lambda_n$.

It is seen that $M_i(x)$ is a polynomial of degree not exceeding i, so that equation (76) belongs to the type considered in *Sets* (pp. 29–31).

THEOREM 22. *The only polynomial solutions of* (76) *are the polynomials* $\{Y_n(x)\}$, *and the only characteristic numbers are therefore* $\{\lambda = \lambda_n\}$.

For let $Y(x)$ be a polynomial satisfying (76) with the value $\lambda = \lambda'$. Then, since the relation between equations (4) and (76) can be traced in *both* directions, the polynomial $y(x)$ given by $y(x) = BEF\{Y(x)\}$ will satisfy (4) for $\lambda = \lambda'$. This can be true only if λ' is one of the numbers λ_n, and in this case we must have $y(x) \equiv y_n(x)$. Hence $Y(x) \equiv Y_n(x)$.

PENNSYLVANIA STATE COLLEGE,
STATE COLLEGE, PA.

ON THE RESULTANT OF A SYSTEM OF FORMS HOMOGENEOUS IN EACH OF SEVERAL SETS OF VARIABLES*

BY

NEAL H. McCOY

Introduction

One of the most fundamental problems in the theory of elimination may be stated as follows. Let

$$(1) \qquad f_i' \qquad (i = 1, 2, \cdots, n)$$

be a set of n general forms homogeneous in the n variables $x_1, x_2, \cdots, x_n$; to determine the polynomial in the coefficients of these forms whose vanishing is a necessary and sufficient condition that the forms (1) simultaneously vanish for a set of values, not all zero, of the variables $x_1, x_2, \cdots, x_n$. This polynomial is called the *resultant* of the system of forms (1). From this standpoint a numerical factor in the resultant is of no consequence though in certain cases it is desirable to introduce some convention as to such a factor.

The most important properties of the resultant of the system (1) are well known and have been obtained by various authors in a variety of ways.† We give a brief account of the method used by König‡ as it is of particular importance in the sequel.

Let us denote by

$$(1') \qquad f_i \qquad (i = 1, 2, \cdots, n)$$

the general non-homogeneous polynomials obtained from (1) by placing one variable, say x_n, equal to unity in each form. We now consider the *module* defined by these polynomials, that is, the system of all polynomials of the form

$$(2) \qquad \phi_1 f_1 + \phi_2 f_2 + \cdots + \phi_n f_n,$$

* Presented to the Society, September 9, 1931; received by the editors May 21, 1932. This paper was practically completed while the author was a National Research Fellow at Princeton University.

† See the Encyklopädie, vol. 1, pp. 260–273; also J. König, *Einleitung in die allgemeine Theorie der algebraischen Grössen*, Leipzig, Teubner, 1903, chapter VI; F. S. Macaulay, *Algebraic Theory of Modular Systems*, Cambridge University Tracts, No. 19, 1916, pp. 3–17; O. Perron, *Algebra* I: *Die Grundlagen*, Göschens Lehrbücherei, vol. 8, 1927.

‡ Op. cit. The exposition given by König is based on earlier work of F. Mertens. For references see König, op. cit., p. 271.

where the ϕ_i are also polynomials in $x_1, x_2, \cdots, x_{n-1}$. It may be shown that there exists one and only one polynomial R in the coefficients of the polynomials (1′) satisfying the following two conditions: (i) R is a member of the module (2), and (ii) R is an irreducible function of the coefficients of these general polynomials. This polynomial R is defined by König to be the resultant of the polynomials (1′) and also of the forms (1). The resultant as thus defined is identical with the polynomial in the coefficients whose vanishing is a necessary and sufficient condition that the forms (1) vanish for a common set of values of the variables. However, this fact is not the center of interest from this point of view. The usual properties of the resultant may be obtained by a method of induction.

It is the purpose of the present paper to consider a certain generalization of the concept of resultant from the point of view of modular systems. Let

$$F_i' \qquad \left(i = 1, 2, \cdots, 1 + \sum_{j=1}^{r} \alpha_j\right) \tag{3}$$

denote a set of general forms homogeneous in the variables of each of $r(\geqq 1)$ sets, there being $\alpha_j + 1$ variables in the jth set $(j = 1, 2, \cdots, r)$ and each $\alpha_j \geqq 1$. Sylvester* seems to have been the first to consider the concept of a resultant of forms of the type (3) and although he did not define the resultant of such a set of forms, he stated without proof a general theorem regarding the degree and weight of the resultant. This theorem is essentially our Theorem 3 (c,d) below.

A definition and a brief discussion of the resultant of the system (3) was given by Lasker† from a point of view somewhat similar to that of the present paper. However, Lasker was not primarily interested in the structure of the resultant but in its use in generalizing certain theorems in the theory of modules and ideals.

Certain special cases have been studied by different authors with, of course, varying points of view. Sylvester and Muir‡ have discussed the resultant of a system of forms linear in each of two sets of variables and have expressed the resultant in the form of a determinant in two or three different

* J. J. Sylvester, *On the degree and weight of the resultant of a multipartite system of equations*, Proceedings of the Royal Society of London, vol. 12 (1862–63), pp. 674–76, or Mathematical Papers, vol. 2, pp. 329–330.

† E. Lasker, *Zur Theorie der Moduln und Ideale*, Mathematische Annalen, vol. 60 (1905), pp. 105–107.

‡ T. Muir, *The resultant of a set of homogeneous lineo-linear equations*, Transactions of the Royal Society of South Africa, vol. 2 (1910–12), pp. 373–380; J. J. Sylvester, *On a question of compound arrangement*, Proceedings of the Royal Society of London, vol. 12 (1862–63), pp. 561–563, or Mathematical Papers, vol. 2, pp. 325–326.

ways. The case of three double binary forms has been considered by Moore and by the present author with the object of expressing the resultant in determinantal form.*

In Part I we give a definition of the resultant of a system of forms of type (3) and deduce some of its fundamental properties. The outline of procedure is essentially that of König† for the classical case $r=1$. Some of his results can be carried over immediately to this more general case and with one or two exceptions we shall refer to König for the proofs wherever possible. However we give in some detail the demonstrations that involve any essential modification or extension.

Part II consists of a generalization of Sylvester's dialytic method of elimination to certain cases of forms of the type here considered. The main result is Theorem 4. As special cases of this theorem we obtain the resultant in the form of a determinant for (i) two ordinary binary forms of arbitrary degrees (Sylvester's determinant); (ii) multiple binary forms of arbitrary degrees in the variables of one set, all the forms being of the same degree in the variables of any other given set; and (iii) forms linear in any number of sets of variables, there being an arbitrary number of variables in each set. The form of the determinant in the third case for two sets of variables is different from the determinants obtained by Muir to which reference was made above.

In general, we obtain more than one determinantal expression for the resultant as the form of the determinant occurring in the statement of Theorem 4 depends in a certain way upon the notation adopted.

I. Definition and fundamental properties of the resultant

1. Notation and preliminary remarks. Let us denote by $x_{j1}, x_{j2}, \cdots, x_{j,\alpha_j+1}$ the variables of the jth set occurring in the forms (3) $(j=1, 2, \cdots, r)$. We shall henceforth let m denote the quantity $1+\sum_{j=1}^{r}\alpha_j$. The degree of F_i^r in the variables of the jth set will be indicated by n_{ij} $(i=1, 2, \cdots, m; =1, 2, \cdots, r)$. We assume throughout that each $n_{ij}>0$; that is, each of the sets of variables actually appears in each form.‡

* T. W. Moore, *Extended results in elimination*, Annals of Mathematics, vol. 30 (1928), pp. 92–100; N. H. McCoy, *On the resultant of three double binary forms*, Ibid., vol. 33 (1932), pp. 177–183. We include here the following additional references which have some relation to the subject of this paper: A. Brill, *Ueber Elimination aus einem gewissen System von Gleichungen*, Mathematische Annalen, vol. 5 (1872), pp. 378–396; T. Muir, *Elimination in the case of equality of fractions whose numerators and denominators are linear functions of the variables*, Transactions of the Royal Society of Edinburgh, vol. 45 (1906), pp. 1–7; K. Th. Vahlen, *Ueber den Grad der Eliminationsresultante eines Gleichungssystems*, Journal für die Reine und Angewandte Mathematik, vol. 113 (1894), pp. 348–352.

† Op. cit. It will be understood henceforth that any reference to this author refers to this book.

‡ A resultant exists under certain conditions even if this restriction is not made. Cf. Lasker, loc. cit., p. 106.

It will be convenient at present to consider in place of the homogeneous forms (3) the general non-homogeneous polynomials

$$F_i \qquad (i = 1, 2, \cdots, m) \tag{4}$$

obtained from them by placing $x_{j,\alpha_j+1}=1$ $(j=1, 2, \cdots, r)$ in each form. By a *general polynomial* we shall mean henceforth a polynomial obtainable in this way from a general form of the type (3).

The totality of variables in all the various sets may be denoted by x, and a will indicate the aggregate of coefficients in all the forms under discussion. Thus $\phi(a, x)$ will represent a polynomial in the coefficients of the set (4) and in certain of the variables.

Let γ_i denote the constant term in F_i, that is, the term containing none of the variables. By $[\phi(a, x)]$ we shall indicate the polynomial obtained from $\phi(a, x)$ by substituting $\gamma_i - F_i$ for γ_i $(i=1, 2, \cdots, m)$.* If we make this substitution only on γ_i $(i=2, 3, \cdots, m)$, we shall indicate the resulting polynomial by $[\phi(a, x)]_1$. It is seen that

$$\phi(a, x) \equiv [\phi(a, x)]_1 + H_2F_2 + \cdots + H_mF_m,$$

where the H's are polynomials. This may be expressed in the usual notation,

$$\phi(a, x) \equiv [\phi(a, x)]_1 \qquad (\text{mod } F_2, F_3, \cdots, F_m). \tag{5}$$

For the sake of completeness we now prove two theorems which are of fundamental importance. The proofs do not differ in any essential from the corresponding proofs in the special case $r=1$ but the second in particular illustrates a method of proof which is important in establishing later theorems.

THEOREM 1.† *If*

$$\phi(a, x) \equiv 0 \qquad (\text{mod } F_1, F_2, \cdots, F_k),$$

then ϕ actually contains all the coefficients occurring in F_1, or

$$\phi(a, x) \equiv 0 \qquad (\text{mod } F_2, F_3, \cdots, F_k).$$

Suppose α is a coefficient in F_1 not occurring in ϕ; then it does not appear in $[\phi(a, x)]_1$. Since

$$\phi(a, x) = H_1F_1 + \cdots + H_kF_k,\ddagger$$

we have

$$[\phi(a, x)]_1 = [H_1]_1F_1.$$

---

* This is the Kronecker substitution.

† König, p. 262.

‡ This relation is of course understood to be an identity in the variables x and the coefficients a. In particular, $\psi(a) = 0$ shall indicate that ψ vanishes identically in the coefficients a.

But $[H_1]_1=0$, as otherwise $[\phi(a, x)]_1$ and consequently $\phi(a, x)$ would contain α. Hence $[\phi(a, x)]_1=0$ and from relation (5) (with m replaced by k) we have the desired result.

THEOREM 2.* *If*

$$\psi(a) \equiv 0 \qquad (\text{mod } F_1, F_2, \cdots, F_k)$$

where $k \leqq m-1$, *then* $\psi(a)=0$.

It is clearly sufficient to prove the theorem for $k=m-1=\sum_{j=1}^{r}\alpha_j$, which is the total number of variables occurring in the polynomials F_i. The theorem is seen to be true in case $m=2$ as in this case we have a single polynomial in a single variable. We accordingly prove the theorem by induction on the total number of variables in our polynomials. We assume the theorem is true for $\mu-1$ general polynomials in $\mu-1$ variables, no matter how the variables are distributed among the various sets.

Let $G_1, G_2, \cdots, G_\mu$ be general polynomials in a total of μ variables, and let ψ denote any polynomial in the coefficients of these polynomials satisfying the relation

$$\psi \equiv 0 \qquad (\text{mod } G_1, G_2, \cdots, G_\mu).$$

By Theorem 1, ψ actually contains the coefficient, say β, of the term x_{11}^b $(b>0)$ in G_1 or

$$\psi \equiv 0 \qquad (\text{mod } G_2, G_3, \cdots, G_\mu).$$

In the latter case we have the identity,

$$\begin{aligned}\psi &= K_2G_2 + \cdots + K_\mu G_\mu \\ &= (K_2)_{x_{11}=0}(G_2)_{x_{11}=0} + \cdots + (K_\mu)_{x_{11}=0}(G_\mu)_{x_{11}=0}.\end{aligned}$$

But $(G_2)_{x_{11}=0}, \cdots, (G_\mu)_{x_{11}=0}$ is a set of $\mu-1$ general polynomials in $\mu-1$ variables, and by the hypothesis of the induction, $\psi=0$. Suppose however that ψ contains β, and when arranged according to powers of β let ψ_s be the coefficient of $\beta^s(s>0)$, the highest power of β occurring in ψ. In the identity

$$\psi = \sum_{i=1}^{\mu} H_iG_i = H_1(\cdots + \beta x_{11}^b + \cdots) + \sum_{i=2}^{\mu} H_iG_i,$$

equate the coefficients of β^s on both sides. We get

$$\psi_s = L_1x_{11}^b + L_2G_2 + \cdots + L_\mu G_\mu,$$

where $L_1=0$ if H_1 is of degree less than $s-1$ in β. Place $x_{11}=0$ and we have

* Cf. König, p. 263.

$$\psi_s = (L_2)_{x_{11}=0}(G_2)_{x_{11}=0} + \cdots + (L_\mu)_{x_{11}=0}(G_\mu)_{x_{11}=0}.$$

By the argument above we find that $\psi_s = 0$, which contradicts our assumption that β actually appeared in ψ. Hence $\psi = 0$.

2. **The fundamental theorem.** We have just shown that there exists no polynomial in the coefficients of the polynomials (4) which belongs to the module defined by $F_1, F_2, \cdots, F_k$ where $k \leqq (m-1)$. That there does exist such a polynomial if $k = m$ is shown by Theorem 3 below. Before stating this theorem we need to give a definition.

By the *weight of a coefficient of F_i with regard to the variables of the kth set*, we shall mean the exponent of x_{k,α_k+1} in the corresponding term of the homogeneous form F_i'.

For convenience let us set

$$L_i = n_{i1}t_1 + n_{i2}t_2 + \cdots + n_{ir}t_r \qquad (i = 1, 2, \cdots, m)$$

where the t's are a set of independent parameters and n_{ij} represents the degree of F_i in the variables $x_{j1}, x_{j2}, \cdots, x_{j,\alpha_j}$, of the jth set. We may now state the following fundamental theorem.

THEOREM 3.* *There exists one and only one*† *rational and integral function, say $R(a)$, of the coefficients of the general polynomials* (4) *with the following properties*:

(a) *$R(a)$ is irreducible*;
(b) $R(a) \equiv 0 \pmod{F_1, F_2, \cdots, F_m}$;
(c) *$R(a)$ is homogeneous and of degree N_i in the coefficients of F_i separately, where*

$$N_i = \textit{coefficient of } t_1^{\alpha_1}t_2^{\alpha_2}\cdots t_r^{\alpha_r} \textit{ in } \prod_{l=1}^{m}{}^{(i)} L_l \ddagger \qquad (i = 1, 2, \cdots, m);$$

(d) *$R(a)$ is isobaric of weight W_k with regard to the variables of the kth set, where*

$$W_k = \textit{coefficient of } t_1^{\alpha_1}\cdots t_{k-1}^{\alpha_{k-1}}t_k^{\alpha_k+1}t_{k+1}^{\alpha_{k+1}}\cdots t_r^{\alpha_r} \textit{ in} \prod_{l=1}^{m} L_l \qquad (k = 1, 2, \cdots, r).$$

This polynomial $R(a)$ is defined to be the *resultant* of the polynomials (4), and is thus defined only to within a numerical factor. Lemma 1 below

* Cf. König, p. 271. Parts (c) and (d) of this theorem were stated by Sylvester, Proceedings of the Royal Society of London, vol. 12 (1862–63), pp. 674–76, or Mathematical Papers, vol. 2, pp. 329–330.

† That is, if $R(a)$ and $R'(a)$ are two polynomials satisfying these conditions, then they differ by only a numerical factor.

‡ This notation indicates, as usual, that l is not to take the value i in this product.

shows that the resultant is determined by the properties (a) and (b) and accordingly the remaining properties must be consequences of these.

The theorem is known to be true in the ordinary case of one set of variables. We have in this case $r=1$, $m=1+\alpha_1$, $N_i=(n_{11}n_{21}\cdots n_{m1})/n_{i1}$, $W_1=n_{11}n_{21}\cdots n_{m1}$. However we shall in the proof of the theorem only make use of this fact for the case of two ordinary polynomials in a single variable, which is the case for $m=2$. *We now assume the theorem for m general polynomials* (4) *where r and α_j $(j=1, 2, \cdots, r)$ are any positive integers such that $m=1+\sum_{j=1}^{r}\alpha_j$.* We shall show that it holds for $m+1$ general polynomials in a total of m variables.

Let

$$G_i \qquad (i=1, 2, \cdots, m+1) \tag{6}$$

be a set of general non-homogeneous polynomials in s sets of variables with β_j variables in the jth set and $m=\sum_{j=1}^{s}\beta_j$. We may without confusion denote the variables of the jth set by $x_{j1}, x_{j2}, \cdots, x_{j,\beta_j}$. Let $\nu_{ij}(>0)$ be the degree of G_i in the variables of the jth set $(i=1, 2, \cdots, m+1; j=1, 2, \cdots, s)$. Further let

$$\overline{L}_i = \nu_{i1}t_1 + \nu_{i2}t_2 + \cdots + \nu_{is}t_s \qquad (i=1, 2, \cdots, m+1),$$

$$\overline{N}_i = \text{coefficient of } t_1^{\beta_1}t_2^{\beta_2}\cdots t_s^{\beta_s} \text{ in } \prod_{l=1}^{m+1}{}^{(i)}\overline{L}_l \qquad (i=1, 2, \cdots, m+1),$$

$$\overline{W}_k = \text{coefficient of } t_1^{\beta_1}\cdots t_{k-1}^{\beta_{k-1}}t_k^{\beta_k+1}t_{k+1}^{\beta_{k+1}}\cdots t_s^{\beta_s} \text{ in } \prod_{l=1}^{m+1}\overline{L}_l \; (k=1, 2, \cdots, s).$$

We wish to show under the hypothesis of the induction that the resultant of the polynomials (6) exists and is of degree $\overline{N}_i$ in the coefficients of G_i and of weight $\overline{W}_k$ with regard to the variables of the kth set.

Before proceeding further we need three lemmas, the first two of which we shall state without proof as they may be readily established as in the case of one set of variables.

LEMMA 1. (König, pp. 267, 272–74.) *If there exists a polynomial ϕ in the coefficients of the general polynomials* (6) *such that*

$$\phi \equiv 0 \qquad (\text{mod } G_1, G_2, \cdots, G_{m+1}),$$

then there exists one and only one irreducible polynomial R' with the same property and ϕ is divisible by R'. Also R' is homogeneous in the coefficients of each polynomial separately and isobaric with regard to each set of variables.

The proof of this lemma does not depend upon the hypothesis of the induction.

Lemma 2. (König, pp. 274–75.) *Let $g(a, x)$ be any polynomial in the coefficients of $F_1, F_2, \cdots, F_m$ and in the variables, of total degree λ in the variables. Then*

$$\left(\frac{\partial R(a)}{\partial \gamma_i}\right)^{\lambda} g(a, x) \equiv h(a) \qquad (\text{mod } F_1, F_2, \cdots, F_m),$$

where $h(a)$ is a polynomial in the coefficients only and γ_i is the constant term in F_i. Further, a necessary and sufficient condition that $g(a, x) \equiv 0$ (mod $F_1, F_2, \cdots, F_m$) is that $h(a) \equiv 0$ (mod $F_1, F_2, \cdots, F_m$).

Lemma 3.* *Let G_i $(i=1, 2, \cdots, m+1)$ be the general polynomials* (6) *and indicate by α the coefficient of the term $x_{11}^{\nu_{11}} x_{21}^{\nu_{12}} \cdots x_{s1}^{\nu_{1s}}$ in G_1. If $\psi(\neq 0)$ is a polynomial in the coefficients of these polynomials and*

$$\psi \equiv 0 \qquad (\text{mod } G_1, G_2, \cdots, G_{m+1}),$$

then in the development of ψ according to powers of α,

$$\psi_\mu = \psi_0 + \cdots + \psi_\mu \alpha^\mu,$$

ψ_μ is divisible by $\overline{R}_1^{\nu_{11}} \overline{R}_2^{\nu_{12}} \cdots \overline{R}_s^{\nu_{1s}}$, where $\overline{R}_k$ is the resultant of the general polynomials, $(G_2)_{x_{k1}=0}, \cdots, (G_{m+1})_{x_{k1}=0}$ $(k=1, 2, \cdots, s)$.

Each $\overline{R}_k$ has the properties of Theorem 3 by the hypothesis of the induction, as it is the resultant of a set of m general polynomials in a total of $m-1$ variables.

We have

$$\psi = H_1 G_1 + H_2 G_2 + \cdots + H_{m+1} G_{m+1}$$

and by equating coefficients of α^μ on both sides† we get

$$\psi_\mu = K_1 x_{11}^{\nu_{11}} x_{21}^{\nu_{12}} \cdots x_{s1}^{\nu_{1s}} + K_2 G_2 + \cdots + K_{m+1} G_{m+1}. \tag{7}$$

Place $x_{k1}=0$ and we have

$$\psi_\mu = (K_2)_{x_{k1}=0} (G_2)_{x_{k1}=0} + \cdots + (K_{m+1})_{x_{k1}=0} (G_{m+1})_{x_{k1}=0},$$

and by Lemma 1, ψ_μ is divisible by $\overline{R}_k$ $(k=1, 2, \cdots, s)$.

Suppose ψ_μ is divisible by $\overline{R}_k^{\nu_{1k}-l}$ but not by $\overline{R}_k^{\nu_{1k}-l+1}$. Then we may write

$$\psi_\mu = \overline{R}_k^{\nu_{1k}-l} \eta, \tag{8}$$

where η does not contain $\overline{R}_k$ as a factor. From (7) we get

$$[\overline{R}_k]_1^{\nu_{1k}-l} [\eta]_1 = K x_{11}^{\nu_{11}} x_{21}^{\nu_{12}} \cdots x_{s1}^{\nu_{1s}}. \tag{9}$$

* This lemma is a generalization of a theorem of Mertens. See König, p. 282. The lemma is stated in unsymmetrical form for convenience of notation.

† The coefficient α actually occurs in ψ by Theorems 1 and 2, hence $\psi_\mu \neq 0$.

It may now be shown that $[\overline{R}_k]_1$ is divisible by x_{k1} but not by x_{k1}^2.* Hence $[\overline{R}]_1^{\nu_{1k}-l}$ is exactly divisible by $x_{k1}^{\nu_{1k}-l}$ and thus by (9), $[\eta]_1$ is divisible by x_{k1}^l. Hence by relation (5) and Lemma 1, η is divisible by $\overline{R}_k$ or $l=0$. But η is not divisible by $\overline{R}_k$, thus $l=0$ and ψ_μ is divisible by $\overline{R}_k^{\nu_{1k}}$. As each of the resultants $\overline{R}_1, \overline{R}_2, \cdots, \overline{R}_s$ contains coefficients not in any of the others and each is irreducible, it follows that ψ_μ is divisible by the required factor.

Let us assume for the moment the existence of the resultant of the general polynomials (6). We can then use the resultant for the ψ of this lemma. Thus the degree of the resultant in the coefficients of $G_i (i=2, 3, \cdots, m+1)$ can not be less than the degree to which these coefficients enter the product $\overline{R}_1^{\nu_{11}}\overline{R}_2^{\nu_{12}} \cdots \overline{R}_s^{\nu_{1s}}$. That is, the degree of the resultant in the coefficients of G_i can not be less than

$$\sum_{j=1}^{s} \nu_{1j}(\text{coefficient of } t_1^{\beta_1} \cdots t_{j-1}^{\beta_{j-1}} t_j^{\beta_j - 1} t_{j+1}^{\beta_{j+1}} \cdots t_s^{\beta_s} \text{ in } \prod_{l=2}^{m+1} {}^{(i)}\overline{L}_l)$$

$$= \text{coefficient of } t_1^{\beta_1} t_2^{\beta_2} \cdots t_s^{\beta_s} \text{ in } \prod_{l=1}^{m+1} {}^{(i)}\overline{L}_l = \overline{N}_i \qquad (i = 2, 3, \cdots, m+1).$$

The polynomial G_1 played an exceptional part in the statement of Lemma 3 but it is clear that any other one could be used in place of G_1. By an argument similar to the above we find that the resultant is of degree not less than $\overline{N}_i$ in the coefficients of G_i $(i=1, 2, \cdots, m+1)$. We proceed to show the existence of the resultant and to show that its degree in the coefficients of G_i is not greater than $\overline{N}_i$.

3. **Existence of the resultant.** It will be convenient to consider two cases according as $\beta_j > 1$ for some j or all $\beta_j = 1$. In the first case we may assume that $\beta_1 > 1$.

Case 1. $\beta_1 > 1$. Let us write the polynomials (6) in the form

$$G_i = \sum A_\lambda^{(i)} x_{11}^{\lambda_1} x_{12}^{\lambda_2} \cdots x_{1\beta_1}^{\lambda_{\beta_1}} \cdots \qquad (i = 1, 2, \cdots, m+1), \tag{10}$$

where the degrees of the variables of the first set are explicitly indicated in each term. For brevity let A indicate the coefficients in this set of polynomials. These polynomials may also be written in the form

$$G_i = \sum B_\lambda^{(i)} x_{11}^{\lambda_1} x_{12}^{\lambda_2} \cdots x_{1,\beta_1-1}^{\lambda_{\beta_1-1}} \cdots \qquad (i = 1, 2, \cdots, m+1) \tag{11}$$

where $B_\lambda^{(i)}$ is a polynomial in $x_{1\beta_1}$ of degree $\nu_{i1} - \sum_{k=1}^{\beta_1-1} \lambda_k$. When written in this form we think of the polynomials as polynomials in s sets of variables, there being $\beta_1 - 1$ in the first set and β_j in the jth set $(j=2, 3, \cdots, s)$. Let B denote the aggregate of coefficients $B_\lambda^{(i)}$ in this set.

* See König, p. 284.

The resultant of $G_1, G_2, \cdots, G_m$ when written in the form (11) exists by the hypothesis of the induction. Let us denote it by $R_{m+1}(B)$ or $R_{m+1}(A, x_{1\beta_1})$ whenever we wish to show that it depends on the coefficients A and also on $x_{1\beta_1}$. It is seen that $R_{m+1}(A, 0)$ is the resultant of the general polynomials

$$(G_1)_{x_1\beta_1=0}, \cdots, (G_2)_{x_1\beta_1=0}.$$

Let us calculate the degree of $R_{m+1}(A, x_{1\beta_1})$ in $x_{1\beta_1}$. The coefficient $B_{\lambda^{(i)}}$ is of degree $\nu_{i1}-\sum_{k=1}^{\beta_1-1}\lambda_k$ in $x_{1\beta_1}$, which is exactly the weight of $B_{\lambda^{(i)}}$ with regard to the β_1-1 variables of the first set. Since this is true for each coefficient and $R_{m+1}(B)$ is isobaric, it is seen by applying Theorem 3 that $x_{1\beta_1}$ occurs in each term of $R_{m+1}(A, x_{1\beta_1})$ to the degree $\overline{N}_{m+1}$. The coefficient of this highest power of $x_{1\beta_1}$ in $R_{m+1}(A, x_{1\beta_1})$ is not zero, as it is the resultant of the general polynomials obtained from $G_1, G_2, \cdots, G_m$, by replacing each $B_{\lambda^{(i)}}$ by the coefficient of the highest power of $x_{1\beta_1}$ occurring in $B_{\lambda^{(i)}}$.

We know that

$$R_{m+1}(A, x_{1\beta_1}) \equiv 0 \qquad (\text{mod } G_1, G_2, \cdots, G_m). \tag{12}$$

Let b_i be the constant term in G_i when written in the form (11), that is, b_i is a polynomial in $x_{1\beta_1}$ but contains no other variables. Denote $\sum_{k=1}^{s}\nu_{m+1,k}$ by p. Then by Lemma 2, we have

$$\left(\frac{\partial R_{m+1}(B)}{\partial b_i}\right)^p G_{m+1} \equiv h(B) \qquad (\text{mod } G_1, G_2, \cdots, G_m). \tag{13}$$

We may also write $h(B)$ as $h(A, x_{1\beta_1})$. Now $h(B)$ is not zero and is of the first degree in the coefficients of G_{m+1}, as $R_{m+1}(B)$ does not contain these coefficients. Suppose $h(A, x_{1\beta_1})$ does not actually contain $x_{1\beta_1}$. Then we have

$$h(A, 0) \equiv 0 \qquad (\text{mod } G_1, G_2, \cdots, G_{m+1}),$$

and by Lemma 1, the resultant of our forms (10) exists and is of at most the first degree in the coefficients of G_{m+1}. But we have shown above that the resultant can not be of degree less than $\overline{N}_{m+1}$ in these coefficients and $\overline{N}_{m+1} \geqq 1$. Hence $\overline{N}_{m+1}$ must in this case be equal to 1 and the resultant is of degree $\overline{N}_{m+1}$ in the coefficients of G_{m+1}.

Suppose then that $h(A, x_{1\beta_1})$ actually contains $x_{1\beta_1}$. It may now be shown that $R_{m+1}(A, x_{1\beta_1})$ and $h(A, x_{1\beta_1})$ have no common factor other than a numerical constant.* Let $S_{m+1}(A)$ be the ordinary resultant of these two as polynomials in $x_{1\beta_1}$. Then $S_{m+1}(A) \neq 0$ and

$$S_{m+1}(A) \equiv 0 \qquad (\text{mod } R_{m+1}(B), h(B)),$$

or by (12) and (13),

* See König, p. 281.

$$S_{m+1}(A) \equiv 0 \qquad (\text{mod } G_1, G_2, \cdots, G_{m+1}).$$

Lemma 1 again establishes the existence of the resultant of the polynomials (10) and we know that $S_{m+1}(A)$ is divisible by this resultant. Since $R_{m+1}(B)$ is of degree $\overline{N}_{m+1}$ in $x_{1\beta_1}$ and $h(A, x_{1\beta_1})$ is linear in the coefficients of G_{m+1}, $S_{m+1}(A)$ is of degree $\overline{N}_{m+1}$ in the coefficients of G_{m+1}. Thus the degree of the resultant in the coefficients of G_{m+1} is not greater than $\overline{N}_{m+1}$, and by the result of §2 this degree is not less than $\overline{N}_{m+1}$. Hence this degree is exactly $\overline{N}_{m+1}$ as we wished to show.

By a similar argument it may be shown that the degree of the resultant in the coefficients of G_i is exactly $\overline{N}_i$ for each i.

Case 2. $\beta_j=1$ $(j=1, 2, \cdots, s)$, $m=s$. There is only one variable in each set; let us denote them by $y_1, y_2, \cdots, y_s$ respectively. The polynomials may be written in the two forms

$$\begin{aligned} G_i &= \sum A_\lambda^{(i)} y_1^{\lambda_1} y_2^{\lambda_2} \cdots y_s^{\lambda_s} \\ &= \sum B_\lambda^{(i)} y_2^{\lambda_2} y_3^{\lambda_3} \cdots y_s^{\lambda_s} \qquad (i = 1, 2, \cdots, m+1), \end{aligned}$$

where $B_\lambda^{(i)}$ is of degree ν_{i1} in y_1.

Let M_i denote the coefficient of $t_2 t_3 \cdots t_s$ in

$$\prod_{j=1}^{m}{}^{(i)}(\nu_{j2}t_2 + \nu_{j3}t_3 + \cdots + \nu_{js}t_s).$$

Now form the resultant of $G_1, G_2, \cdots, G_m$ as polynomials in $y_2, y_3, \cdots, y_s$, say $R_{m+1}'(B)=R_{m+1}'(A, y_1)$. Then by the hypothesis of the induction, $R_{m+1}'(A, y_1)$ is of degree

$$M = \nu_{11}M_1 + \nu_{21}M_2 + \cdots + \nu_{m1}M_m$$

in y_1. But M is the coefficient of $t_1 t_2 \cdots t_s$ in

$$\prod_{j=1}^{m}(\nu_{j1}t_1 + \nu_{j2}t_2 + \cdots + \nu_{js}t_s),$$

and thus $M=\overline{N}_{m+1}$. From this point we may proceed exactly as in the previous case and the details will be omitted.

Before considering the proof of part (d) of Theorem 3 it is desirable to pass back to homogeneous forms.

4. **The resultant of homogeneous forms.** Let us make our general polynomials (6) homogeneous in each of the s sets of variables by introducing new variables, $z_j=x_{j,\beta_j+1}$ $(j=1, 2, \cdots, s)$. These homogeneous forms will be denoted by

$$G_i' \qquad (i = 1, 2, \cdots, m+1). \tag{14}$$

Let R indicate now the resultant of the polynomials (6). Then we have

$$R = \sum_{i=1}^{m+1} H_i G_i.$$

This goes over in the homogeneous case to

$$(15) \qquad R z_1^{\rho_1} z_2^{\rho_2} \cdots z_s^{\rho_s} = \sum_{i=1}^{m+1} H_i' G_i',$$

where ρ_k is the weight of R with regard to the variables of the kth set. We show below that $\rho_k = \overline{W}_k$ but for the present its value is immaterial.

In G_i' let us place $x_{1j_1} = x_{2j_2} = \cdots = x_{sj_s} = 1$, where these are any variables of the respective sets, and denote the resulting non-homogeneous polynomials by $G_i^{(j)}$. Then from (15) we have

$$R z_1^{\rho_1} z_2^{\rho_2} \cdots z_s^{\rho_s} = \sum_{i=1}^{m+1} H_i^{(j)} G_i^{(j)}.$$

By the Kronecker substitution we see that $[R] = 0$ and thus

$$R \equiv 0 \qquad (\text{mod } G_1^{(j)}, G_2^{(j)}, \cdots, G_{m+1}^{(j)}).$$

If we denote by $R^{(j)}$ the resultant of the polynomials $G_i^{(j)}$, then this relation shows that R is divisible by $R^{(j)}$ and by reversing the process we see that $R^{(j)}$ is divisible by R. Thus we get the same resultant, defined only to within a numerical factor, no matter which one of the variables of each set in G_i we place equal to unity. The resultant R of G_i $(i = 1, 2, \cdots, m+1)$ we accordingly define to be the *resultant of the homogeneous forms* (14).

Let $\overline{G}_i'$ $(i = 1, 2, \cdots, m+1)$ denote the set of forms (14) after we have made a general non-singular linear transformation on the variables of say the first set, and let $\overline{R}$ be the resultant of this transformed set of forms. It follows immediately* that

$$(16) \qquad \overline{R} = RU,$$

where U is a form in the coefficients of the transformation only. As a matter of fact U must be a power of the determinant of the transformation.† As a special case we see that the resultant is unchanged if the variables are permuted in any way within the set in which they occur.

We now prove the following:

* König, p. 293.

† See Bôcher, *Higher Algebra*, p. 220. Thus the resultant is an invariant under independent linear transformations of the various sets.

LEMMA 4. *Let $x_{1j_1}, x_{2j_2}, \cdots, x_{sj_s}$ be any variables of the respective sets in G_i'. Then the resultant of the forms G_i' contains the coefficient of*

$$x_{1j_1}^{\nu_{k1}}\, x_{2j_2}^{\nu_{k2}} \cdots x_{sj_s}^{\nu_{ks}}$$

in G_k' to the degree $\overline{N}_k$.

In view of the previous remarks it will be sufficient to prove this lemma for the case where $j_r=\beta_r+1$ $(r=1, 2, \cdots, s)$. For convenience of notation let us consider the case $k=2$. Let α_2 denote the coefficient under consideration, that is, α_2 is the constant term in G_2. By Lemma 3, α_2 enters the resultant to a degree as great as the degree to which it enters $\overline{R}_1^{\nu_{11}}\,\overline{R}_s^{\nu_{12}}\cdots\overline{R}_s^{\nu_{1s}}$, which by the use of induction on the number of variables is

$$\sum_{k=1}^{s}\nu_{1k}\left[\text{coefficient of } t_1^{\beta_1}\cdots t_{k-1}^{\beta_{k-1}}t_k^{\beta_k-1}t_{k+1}^{\beta_{k+1}}\cdots t_s^{\beta_s} \text{ in } \prod_{i=3}^{m+1}(\nu_{i1}t_1+\cdots+\nu_{is}t_s)\right].$$

But this is $\overline{N}_2$. The proof is unchanged for any $k\neq 1$. If $k=1$, we need only to change the way in which Lemma 3 has been stated for simplicity of notation.

It now follows from Lemma 3 by a consideration of the degrees that the resultant of the forms (14) contains the term

$$\alpha^{\overline{N}_1}\,\overline{R}_1^{\nu_{11}}\,\overline{R}_2^{\nu_{12}}\cdots\overline{R}_s^{\nu_{1s}} \tag{17}$$

with at most a numerical coefficient. Here α is the coefficient of $x_{11}^{\nu_{11}}x_{21}^{\nu_{12}}\cdots x_{s1}^{\nu_{1s}}$ in G_1. We may determine the weight of this expression (17) with regard to the kth set of variables by the hypothesis of the induction and it is found to be $\overline{W}_k$. Since by Lemma 1 the resultant is isobaric, each term of the resultant is of weight $\overline{W}_k$ with regard to the kth set of variables. This proves part (d) of Theorem 3 and thus completes the proof of the theorem.

II. THE RESULTANT IN DETERMINANT FORM

5. We now pass to the problem of expressing the resultant in determinant form in certain special cases.

Let

$$\phi_i \qquad (i = 1, 2, \cdots, m) \tag{18}$$

be a set of general forms homogeneous in each of $s+t=r$ $(r\geqq 1,\ s,\ t\geqq 0)$ sets of variables, there being α_j+1 variables in the jth set $(\alpha_j\geqq 1, j=1, 2, \cdots, r)$. We assume further that $\alpha_1=\alpha_2=\cdots=\alpha_s=1$, and hence $m=s+1+\sum_{s+1}^{r}\alpha_j$. Also the degrees of these forms in the various sets are assumed to be those given in the following table:

		(1)	(2)	(3) $\cdots$ (s)	$(s+1)$	$(s+2)$	$\cdots(s+t=r)$
	ϕ_1	n_{11}	n_2	$n_3 \cdots n_s$	1	1 $\cdots$	1
(19)	ϕ_2	n_{21}	n_2	$n_3 \cdots n_s$	1	1 $\cdots$	1
	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$
	ϕ_m	n_{m1}	n_2	$n_3 \cdots n_s$	1	1 $\cdots$	1

Here the degree of ϕ_i in the variables of the jth set is found at the intersection of the ith row and jth column. The numbers n_{j1} $(j=1, 2, \cdots, m)$, n_2, n_3, $\cdots$, n_s are arbitrary positive integers.

We shall express in determinant form the resultant of this system of forms. As a special case if $t=0$, we have a set of multiple binary forms, the degree in the variables of the kth set being the same for each form if $k>1$. If, further, $r=1$, we have two ordinary binary forms of arbitrary degrees and our form of the resultant reduces to the Sylvester determinant. If on the other hand $s=0$, we have a set of forms linear in each of t sets of variables, there being an arbitrary number of variables in each set.

We now state the principal result of this section.

THEOREM 4. *Let ϕ_i $(i=1, 2, \cdots, m)$ be the general forms* (18) *and consider all possible equations of the type*

$$\rho_i \phi_i = 0 \qquad (i = 1, 2, \cdots, m). \tag{20}$$

Here ρ_i represents a power product of the variables of such a degree that $\rho_i\phi_i$ is homogeneous of degree $\sum_{i=1}^{m} n_{i1}-1$ in the variables of the first set; of degree $(m-k+1)n_k-1$ in the variables of the kth set $(k=2, 3, \cdots, s)$; of degree $\alpha_{s+l+1}+ \cdots +\alpha_r+1$ in the variables of the $(s+l)$th set $(l=1, 2, \cdots, t-1)$; and of the first degree in the variables of the rth set. Considering these power products of the variables as unknowns, we have in the set* (20) *the same number of equations as unknowns and the determinant of the coefficients of the unknowns is the resultant of the given forms.*

Let us calculate, for example, the number of the equations (20) arising from ϕ_1. Making use of the fact that the number of terms in a general polynomial, homogeneous of degree N in one set of M variables, is

$$\binom{M+N-1}{N} = \frac{(M+N-1)!}{N!(M-1)!},$$

we see that the number of equations arising from ϕ_1 is

* That is, ρ_i does not contain the variables of the rth set if $t > 0$.

$$\left(\sum_{i=2}^{m} n_{i1}\right)(m-2)(m-3)\cdots(m-s)n_2n_3\cdots n_s\binom{\alpha_{s+1}+\alpha_{s+2}\cdots+\alpha_r}{\alpha_{s+1}}$$

$$\times\binom{\alpha_{s+2}+\cdots+\alpha_r}{\alpha_{s+2}}\cdots\binom{\alpha_{r-1}+\alpha_r}{\alpha_{r-1}}.$$

Remembering that $m=s+1+\sum_{s+1}^{r}\alpha_j$, this may be reduced to the form

$$\left(\sum_{i=2}^{m} n_{i1}\right)n_2n_3\cdots n_s[(m-2)!]/(\alpha_{s+1}!\alpha_{s+2}!\cdots\alpha_r!). \tag{21}$$

The total number of the equations (20) is found to be

$$\left(\sum_{i=1}^{m} n_{i1}\right)n_2n_3\cdots n_s[(m-1)!]/(\alpha_{s+1}!\alpha_{s+2}!\cdots\alpha_r!), \tag{22}$$

and a calculation similar to the above shows that this is also the total number of unknowns. We thus have the same number of equations as unknowns.

Let D be the determinant of the coefficients of the unknowns in the equations (20). Assume for the present that $D\not\equiv 0$ for general forms (18) which we are considering. Let $\beta_1, \beta_2, \cdots, \beta_r$ denote the degrees of the power products in the several sets of variables in equations (20) and suppose that the elements of the last column of D are the coefficients of

$$x_{12}^{\beta_1}x_{22}^{\beta_2}\cdots x_{s2}^{\beta_s}\cdots x_{r,\alpha r+1}^{\beta_r}$$

in these equations. Multiply each column of D by the power product of which its elements are coefficients and add to the last column. Each element of the last column is now of the form $\rho_i\phi_i$. Hence if we expand D in terms of the elements of the last column we get

$$Dx_{12}^{\beta_1}x_{22}^{\beta_2}\cdots x_{s2}^{\beta_s}\cdots x_{r,\alpha r+1}^{\beta_r}\equiv 0 \quad (\mathrm{mod}\ \phi_1, \phi_2, \cdots, \phi_m).$$

From the discussion in §4 and Lemma 1 it follows that D is divisible by the resultant of the given forms.

Now D is clearly homogeneous of degree given by the expression (21) in the coefficients of ϕ_1. Let us calculate by Theorem 3 the degree of the resultant in these coefficients. This degree is the coefficient of $t_1t_2\cdots t_s \cdot t_{s+1}^{\alpha_{s+1}}\cdots t_r^{\alpha_r}$ in

$$\prod_{i=2}^{m}(n_{i1}t_1+n_2t_2+\cdots+n_st_s+t_{s+1}+\cdots+t_r).$$

But this is

$$\left(\sum_{i=2}^{m} n_{i1}\right)\Big[\text{coefficient of } t_2 \cdots t_s t_{s+1}^{\alpha_{s+1}} \cdots t_r^{\alpha_r} \text{ in}$$

$$(n_2 t_2 + \cdots + n_s t_s + t_{s+1} + \cdots + t_r)^{m-2}\Big]$$

$$= \left(\sum_{i=2}^{m} n_{i1}\right) n_2 n_3 \cdots n_s[(m-2)!]/(\alpha_{s+1}!\alpha_{s+2}! \cdots \alpha_r!).$$

Thus the degree of D in the coefficients of ϕ_1 is the same as the degree of the resultant in these coefficients and similarly for each $\phi_i (i=2, 3, \cdots, m)$. As D contains the resultant as a factor, D must be the resultant provided $D \not\equiv 0$.

We proceed to show that $D \not\equiv 0$ by a process of induction. We assume Theorem 4 for the case of the proper number of forms of the general type (18) in fewer variables. It is known to be true for the case of two ordinary binary forms.

Let ω_i $(i=1, 2, \cdots)$ represent the power products of the variables occurring in the equations (20), that is, the power products of the degrees mentioned in the statement of Theorem 4. We first of all specialize ϕ_1 by placing

$$\phi_1 = \phi_1' = x_{11}^{n_{11}} x_{21}^{n_2} \cdots x_{s1}^{n_s} x_{s+1,1} \cdots x_{r1}. \tag{23}$$

Then in each row of D arising from ϕ_1' we have one and only one element different from zero and it is equal to unity. The columns of D in which a 1 thus occurs are those whose elements are coefficients of an ω_i which is divisible by ϕ_1'. Let us strike these rows and columns from D and denote the remaining determinant by D'. Thus $D' = \pm D$.

The power products ω_i not divisible by ϕ_1' may be arranged in mutually exclusive sets as follows. Let $\omega_1^{(q)}$ denote those which contain x_{11} exactly to the degree q $(q=0, 1, \cdots, n_{11}-1)$; $\omega_p^{(q)}$ those divisible by $x_{11}^{n_{11}} x_{21}^{n_2} \cdots x_{p-1}^{n_{p-1}}$ and containing x_{p1} to the degree q $(q=0, 1, \cdots, n_p-1; p=2, 3, \cdots, s)$; ω_{s+k} those divisible by $x_{11}^{n_{11}} x_{21}^{n_2} \cdots x_{s1}^{n_s} x_{s+1,1} \cdots x_{s+k-1,1}$ but not divisible by $x_{s+k,1}$ $(k=1, 2, \cdots, r-s)$. Each power product ω_i falls into one and only one of these sets or is divisible by ϕ_1'.

Consider now the set of all power products ρ multiplying $\phi_2, \cdots, \phi_m$, in equations (20). These are of various degrees. The same power product may occur multiplying different forms; in this case we count it as many times as it appears. We suppose further that these ρ's are so labeled that having given a particular ρ we know which form it multiplies in the equations (20). Thus specifying a given ρ designates a row of D'. We now define $\rho_1^{(q)}$, $\rho_p^{(q)}$ and ρ_{s+k} by the same conditions used in defining $\omega_1^{(q)}$, $\omega_p^{(q)}$ and ω_{s+k} respectively.

In particular $\rho_1^{(q)}$ is the set of ρ's which contain x_{11} to the degree q $(q=0, 1, \cdots, n_{11}-1)$, and so on. To get this set we first select those multipliers of ϕ_2 with this property, then those multiplying ϕ_3, and so on, each power product being taken as many times as it appears. Each ρ falls into one and only one of the above sets and none of them is divisible by ϕ_1'.

A direct calculation shows that the number of elements in the sets $\rho_1^{(q)}$, $\rho_p^{(q)}$, ρ_{s+k} is exactly the number of power products ω in the sets $\omega_1^{(q)}$, $\omega_p^{(q)}$ and ω_{s+k} respectively. By a proper arrangement of rows and columns we may therefore write $\pm D'$ in the form

$$
\begin{array}{c|cccccc|}
 & \omega_1^{(0)} & \omega_1^{(1)} & \cdots & \omega_1^{(n_{11}-1)} & \cdots & \omega_r \\
\hline
\rho_1^{(0)} & \boxed{D_1^{(0)}} & & & & & \\
\rho_1^{(1)} & & \boxed{D_1^{(1)}} & & & & \\
\vdots & & & \ddots & & & \\
\rho_1^{(n_{11}-1)} & & & & \boxed{D_1^{(n_{11}-1)}} & & \\
\vdots & & & & & \ddots & \\
\rho_r & & & & & & \boxed{D_r}
\end{array}\,.
$$

In this arrangement, the elements falling in the square array $D_1^{(0)}$ are those in a row of D' denoted by a ρ of the set $\rho_1^{(0)}$ and in a column designated by an ω of the set $\omega_1^{(0)}$, and so on. We suppose the order of the sets of ω's from left to right is

$$\omega_1^{(0)}, \cdots, \omega_1^{(n_{11}-1)}, \omega_2^{(0)}, \cdots, \omega_2^{(n_2-1)}, \cdots, \omega_s^{(n_s-1)}, \omega_{s+1}, \cdots, \omega_r.$$

The ρ's are arranged in the same order from top to bottom. Now there is no non-zero element vertically below any of the square arrays D. For example, consider the set of columns $\omega_p^{(q)}$ $(p>1)$. Every set of ρ's following $\rho_p^{(q)}$ contains

$$x_{11}^{n_{11}}\, x_{21}^{n_2} \cdots x_{p-1,1}^{n_{p-1}} x_{p1}^{q+1}$$

as a factor, while $\omega_p^{(q)}$ does not contain this term as a factor. This observation is all that is necessary to obtain this result.

Thus we have

$$D' = \pm D_1^{(0)} \cdots D_1^{(n_{11}-1)} \cdots D_r,$$

where these denote the determinants of the arrays indicated.

Let S_k denote the resultant of

$$(\phi_2)_{x_{k1}=0}, \cdots, (\phi_m)_{x_{k1}=0} \qquad (k = 1, 2, \cdots, r)$$

with the understanding that if $\alpha_k=1$, as is certainly the case for $k=1, 2, \cdots, s$, we also place $x_{k2}=1$. It may now be shown that

$$\begin{aligned} D_1^{(q)} &= S_1 & (q &= 0, 1, \cdots, n_{11}-1), \\ D_p^{(q)} &= S_p & (p &= 2, 3, \cdots, s;\ q = 0, 1, \cdots, n_p - 1), \\ D_{s+k} &= S_{s+k} & (k &= 1, 2, \cdots, r-s). \end{aligned} \tag{24}$$

We make the calculation for a typical case, say $D_s^{(q)}$ for convenience. Denote by $\bar{\phi}_2, \cdots, \bar{\phi}_m$ the general forms obtained from $\phi_2, \cdots, \phi_m$ by placing $x_{s1}=0$, $x_{s2}=1$ in each form. Then apply Theorem 4, as it is true by the hypothesis of the induction. We shall use the notation as above, for example the rth set of variables will denote those variables which belonged to the rth set in the forms (18) although it is only the $(r-1)$st set here, as the sth set is lacking. We have then equations of the type

$$\eta_i\bar{\phi}_i = 0 \qquad (i = 2, 3, \cdots, m), \tag{25}$$

where $\eta_i\bar{\phi}_i$ is of degree $\sum_{i=2}^{m} n_{i1}-1$ in the variables of the first set; of degree $(m-k)n_k-1$ in the variables of the kth set $(k=2, 3, \cdots, s-1)$; of degree $\alpha_{s+l+1}+ \cdots +\alpha_r+1$ in the variables of the $(s+l)$th set $(l=1, 2, \cdots, t-1)$; and of the first degree in the variables of the rth set. But we obtain exactly these power products occurring in the equations (25) if we divide those of the set $\omega_s^{(q)}$ by their common factor,

$$x_{11}^{n_{11}}\, x_{21}^{n_2} \cdots x_{s-1,1}^{n_{s-1}}\, x_{s1}^{q}\, x_{s2}^{n_s-q}.$$

Similarly we obtain all these power products η_i in (25) by dividing those of the set $\rho_p^{(q)}$ by this same factor.

Now the determinant of the coefficients of the unknowns in the equations (25) is S_s. Multiply these equations by the common factor of the elements of $\omega_s^{(q)}$, and we have the equivalent system of equations

$$\rho_s^{(q)}\bar{\phi}_i = 0 \qquad (i = 2, 3, \cdots, m).$$

The determinant of the coefficients in these equations is $D_s^{(q)}$ as the power products in this set of equations are exactly those of $\omega^{(q)}$, and $D_s^{(q)}$ is seen to contain no coefficient not occurring in the forms $\bar{\phi}_2, \cdots, \bar{\phi}_m$. Thus we see that $D_s^{(q)} = S_s$. In a like manner the other relations (24) may be verified. We have then that

$$D' = S_1^{n_{11}} S_2^{n_2} \cdots S_s^{n_s} S_{s+1} \cdots S_r,$$

and no one of these factors is zero. Thus for general forms D is not zero. This completes the proof of Theorem 4.

The form of the determinant D obtained for a given system of forms (18) clearly depends upon the convention as to which set of variables is the second, which the third, and so on. Thus in general we have a variety of determinantal expressions for the resultant of forms of the type (18).

Smith College,
Northampton, Mass.

AN AXIOMATIC BASIS FOR PLANE GEOMETRY*

BY
STEWART S. CAIRNS

1. **The axioms.** The fourth appendix of Hilbert's *Grundlagen der Geometrie*† is devoted to the foundation of plane geometry on three axioms pertaining to transformations of the plane into itself. The object of the present paper is to attain the same end by quicker and simpler means. The simplifications are made possible by using orientation-reversing transformations‡ and changing Hilbert's second and third axioms.

The (x, y)-plane will mean the set of all distinct ordered pairs of real numbers. The terms of analytic geometry, to which no geometric content need be given, will be used, modified by the prefix (x, y) where ambiguity might arise. Thus we shall refer to (x, y)-*distance*, (x, y)-*lines*, and so on.

The *general plane*, p, will be any set of objects, called *points*, which can be put in one-to-one correspondence with the points of the (x, y)-plane. For convenience, we shall speak of the points of p as if they were identical with their images under such a correspondence.

The following axioms pertain to a set, T, of continuous§ one-to-one transformations of p into itself. A transformation of the set which leaves two distinct points fixed and reverses orientation will be called a *reflection*.

Axiom 1. *The transformations T form a group.*

Axiom 2. *If A and B are two points of p, T contains a reflection leaving A and B fixed.*

Axiom 3. *Let T_A denote the subset of T containing all the transformations thereof which leave A fixed. If T_A contains transformations carrying pairs of points arbitrarily near a given pair of points (B, C) into an arbitrarily small neighborhood of a pair (D, E), then T_A contains a transformation carrying (B, C) into (D, E).*

2. **The curve γ.** Our first object is to establish the following theorem, which, like Lemma 1 below, is similar to a result employed by Hilbert (loc. cit.).

* Presented to the Society, September 9, 1930; received by the editors July 10, 1932. This paper was partly written while the writer was studying under Professor B. von Kerékjártó at Szeged University as a Travelling Fellow from Harvard University.

† D. Hilbert, *Grundlagen der Geometrie*, 1930, pp. 178–230.

‡ Suggested by Hilbert, loc. cit., p. 182.

§ That is, continuous in terms of (x, y)-distance.

THEOREM 1. *Every neighborhood of A contains a simple closed curve, γ, enclosing A and preserved by each of the transformations T_A.*

We shall assume the Jordan separation theorem and the following converse thereof:

(A) *A locally connected* set of points which divides the (x, y)-plane into two regions, one of them finite, and forms their common boundary, is a simple closed curve.*†

LEMMA 1. *For any given positive ϵ, there is a neighborhood, N, of A on p, no point of which is carried to a distance ϵ from A by any of the transformations T_A (Axiom 3).*

Otherwise, let P_i $(i=1, 2, \cdots)$ be a point within distance $\epsilon/2^i$ of A, whose image, Q_i, under one of the transformations T_A is at distance ϵ from A. Then T_A contains transformations carrying points arbitrarily near A into an arbitrarily small neighborhood of any cluster point, Q, of $(Q_1, Q_2, \cdots)$. Therefore, by Axiom 3, T_A contains a transformation carrying A into Q. But this is impossible, for the transformations T_A all leave A fixed.

LEMMA 2. *Let c be any simple closed curve on N enclosing A. The set, Γ, of all points into which points on c are carried by the transformations T_A is closed. The transformations T_A all preserve Γ.*

Any cluster point, P, of Γ is limit of some series $(P_1, P_2, \cdots)$ on Γ. By definition of Γ, one of the transformations T_A carries a point Q_i $(i=1, 2, \cdots)$ on c into P_i. Hence, if Q is a cluster point, necessarily on c, of $(Q_1, Q_2, \cdots)$, T_A (see Axiom 3) contains transformations carrying points arbitrarily near Q into an arbitrary neighborhood of P. Therefore (Axiom 3), T_A contains a transformation carrying Q into P. Hence P is on Γ, and Γ is closed.

Consider the image, P', of any point, P, on Γ under any transformation, T_1, of the set T_A. Let T_0 be one of the transformations T_A carrying some point, Q, on c into P. Then T_0T_1 carries Q into P'. Since T_0T_1 leaves A fixed and belongs to T (Axiom 1), it belongs also to T_A. Therefore P' is on Γ. This completes the proof.

---

* A point set, S, is said to be *locally connected* if, for any $\epsilon>0$ and any point P, of S, there exists a positive distance, δ, such that all points of S within distance δ of P are connected with P by a subset of S entirely within distance ϵ of P.

† Essentially in this form, the theorem is given by J. R. Kline, these Transactions, vol. 21 (1920), p. 452. It is a ready consequence of Hahn's characterisation of continuous curves, Jahresbericht der Deutschen Mathematiker Vereinigung, vol. 23 (1914), p. 318, together with a theorem by R. L. Moore, Bulletin of the American Mathematical Society, vol. 23 (1917), p. 233, that any two points of a continuous curve, S, can be joined on S by a simple Jordan arc.

LEMMA 3. *The complement of* Γ *on the* (x, y)*-plane contains just one unlimited region,* R*. The boundary,* γ*, of* R *divides the* (x, y)*-plane into two regions,* R *and* R_0*, and forms their common boundary.*

The first part of the lemma follows from Lemmas 1 and 2. It also follows from these lemmas that γ is on Γ. Let P denote any point neither in R nor on γ, and k a simple arc through P with just its end points, P_1 and P_2, on γ. Some transformation, T_j $(j=1, 2)$, of the set T_A carries a point, Q_j, of c into P_j (Lemma 2). A simple arc inside c joining A to Q_j is carried by T_j into an arc, k_j, joining A to P_j but not meeting either R or γ. Because R is connected, $(k+k_1+k_2)$ cannot enclose any point of either R or its boundary, γ. Therefore γ cannot separate P from A. Hence all points neither in R nor on γ are in a single region, R_0. By such a curve as k_1, any point on γ can be joined to A inside R_0. Therefore, γ is the common boundary of R_0 and R_1.

LEMMA 4. *The boundary* γ *is locally connected.*

Suppose that at some point, X, γ is not locally connected. Then a positive number, d, exists, so small that every neighborhood of X contains points on γ not connected with X by any continuum on γ entirely within distance d of X. Let P_i $(i=1, 2, \cdots)$ be one such point within distance $d/2^i$ of X. Let C denote the (x, y)-circle of radius d about X and K_i the set of all points connected with P_i on γ inside C. Then, if K_i and K_j have a point in common, they coincide. It may readily be seen that K_i contains all its cluster points inside C. Therefore, at most a finite number of the K's can coincide with any one of them. Otherwise, an infinite subset of $(P_1, P_2, \cdots)$ would belong to one of the K's, which would therefore contain X and join it in C to certain of the P's. Hence, with no loss of generality, we may assume that the K's are all distinct.*

Let C_1 be the circle of radius $d/2$ about X, and K_i' $(i=1, 2, \cdots)$ a closed connected subset of K_i joining P_i to C_1 but containing no points outside C_1. Let C_2 be the circle of radius $d/4$ about X. Without loss of generality, we assume* that K_i' contains a point, Q_i, on C_1 and a point, S_i, on C_2 such that $(Q_1, Q_2, \cdots)$ converges to a limit, Q, monotonically on the arc Q_1Q_2Q, and $(S_1, S_2, \cdots)$ converges to a limit, S, monotonically on the arc S_1S_2S. Consider, for any $i>1$, a simple closed curve made up of two arcs, k_1 and k_2, joining A to S_i inside R_0 (Lemma 3, proof). Suppose k_1 meets the arc $\alpha_i \equiv Q_{i-1}Q_iQ_{i+1}$ on C_1, but not the broken line $\beta_i \equiv P_{i-1}P_iP_{i+1}$, whereas k_2 meets β_i but not α_i. Then (k_1+k_2) separates S_{i-1} from S_{i+1}. For, let a simple

* To avoid excessive notation, we assume for the K's several properties enjoyed by some infinite subset thereof.

closed curve, γ_0, be formed by adding to α_i and β_i a pair of arcs joining Q_{i-1} to P_{i-1} and Q_{i+1} to P_{i+1}, respectively. Let these latter curves pass through S_{i-1} and S_{i+1}, respectively, and lie so near to K'_{i-1} and K'_{i+1} that γ_0 encloses S_i and is met by k_1 only on α_i and by k_2 only on β_i. Then (k_1+k_2) clearly contains just one arc inside γ_0 separating S_{i-1} from S_{i+1}. Therefore S_{i-1} and S_{i+1} are separated by the closed curve (k_1+k_2). But this is impossible, for no curve in R_0 can enclose points of R (Lemma 3). Therefore, if c_i is a simple closed curve in R_0 through A and S_i, then α_i (or β_i) meets both of the arcs into which c_i is divided by A and S_i.

Now S_i $(i=1, 2, \cdots)$, being on Γ (Lemmas 2, 3), is image, under a transformation, T_i, of the set T_A, of some point, E_i, on the curve c of Lemma 2. Let c' be a simple closed curve through A which contains no points outside c but has in common with c an arc through a cluster point, E, of $(E_1, E_2, \cdots)$. With no loss of generality, we assume* that all the E's lie on c' and that $(E_1, E_2, \cdots)$ converges to E. Then T_i carries c' into a simple closed curve, c_i, to which the conclusion of the preceding paragraph applies.† We shall treat only the case where both the arcs AS_i $(i=1, 2, \cdots)$ on c_i meet α_i.‡ In this case, c' passes through two points, E_i' and E_i'', separated on c' by (A, E_i), where the images of (E_i', E_i'') under T_i are on α_i. Without loss of generality, we assume* that $(E_1', E_2', \cdots)$ and $(E_1'', E_2'', \cdots)$ converge to a pair of points, E' and E'', respectively. Now, by definition, α_i, for i large enough, is in an arbitrarily small given neighborhood of Q. Hence, since T_i carries (E_i', E_i'') onto α_i, T_A contains a transformation (Axiom 3) carrying (E', E'') into Q. Hence $E'\equiv E''$. Since A and E_i separate E_i' and E_i'' on c', E' and E'' can coincide only at A or at E. But A cannot go into Q under any of the transformations T_A. Hence $E'\equiv E''\equiv E$. Then, for i large enough, T_i carries a pair of points (E_i, E_i') arbitrarily near E into an arbitrary neighborhood of the pair (S, Q). Therefore (Axiom 3) T_A contains a transformation carrying E' into (S, Q). This contradicts the one-to-one-ness of the transformations and establishes the lemma.

Theorem 1 *above is an immediate consequence of* (A) *together with Lemmas* 2, 3, *and* 4.

3. **Lines and reflections.** The set of all fixed points under a reflection (see §1) will be called a *line*.

---

* See footnote on p. 236.

† To show this, c' may be slightly deformed, if necessary, so that its image meets γ only at S_i.

‡ The method applies equally well if both arcs meet β_i. We need only replace α by β and Q by P. In assuming that the arc AS_i meets α_i (or β_i) for all values i, we employ the convention stated in the footnote on p. 236.

Lemma 1. *An orientation-preserving transformation, τ, of the group T which preserves a simple closed curve, γ, and leaves one point of γ fixed, leaves every point of γ fixed.**

If A denote the known fixed point on γ, then τ and all its powers belong to the set T_A (Axiom 3). Let P_i be the image of P_0 under the ith power of τ. Ascribing a positive sense to γ, consider the arc $\overrightarrow{AP_0}$. If it passes through P_1, then the arc $\overrightarrow{AP_1}$, being the image under τ of $\overrightarrow{AP_0}$, passes through P_2; and, in general, $\overrightarrow{AP_i}$ $(i=2, 3, \cdots)$ passes through P_{i+1}, but not P_{i-1}. On the other hand, if $\overrightarrow{P_0A}$ on γ passes through P_1, then $\overrightarrow{P_iA}$ contains P_{i+1} but not P_{i-1}. Thus, in either case, $(P_1, P_2, \cdots)$ is a monotonic series on the curve γ. If P is its limit, then, for i sufficiently large, the two points (P_i, P_{i+1}) are in an arbitrary preassigned neighborhood of P. But these two points are images, under τ^i, of (P_0, P_1). Therefore (Axiom 3) some transformation of the set T_A carries both P_0 and P_1 into P. This implies that P_0 and P_1 coincide, and hence that every point of C is fixed under τ.

Corollary. *The transformation τ is the identity.*

First, since τ is continuous, the set, S, of its fixed points is closed with respect to p. Suppose S does not coincide with p, and consider the largest connected subset, S_1, of S which contains C. Let k be a simple curve joining a point, Q, of $(p-S)$ to a point, B, of S_1, but not containing any point of (S_1-B). Let γ' be a simple closed curve about B which meets both k and S_1 and which is carried into itself by τ (Theorem 1). Since γ' meets S_1, all its points are fixed under τ and hence belong to S_1. This contradicts the definition of k and thus establishes the corollary.

Lemma 2. *A simple closed curve which is preserved by a reflection, ρ, is met in just two points by the line which ρ defines.*

For, an orientation-reversing transformation which preserves a simple closed curve leaves just two points on the curve fixed.

Lemma 3. *The identity is the only orientation-preserving transformation of the group T which leaves every point of a line, L, fixed.*

This follows from the preceding results of this section applied to the curve γ of Theorem 1, where A is on L.

Theorem 2. *A reflection, ρ, is involutory. No two different reflections define the same line, L.*

---

* The proof is patterned after one by Hilbert, loc. cit., pp. 204, 205.

By Lemma 3, $\rho\rho$ is the identity. Also, if ρ and ρ' both define L, $\rho\rho'$ is the identity. Therefore $\rho = \rho'$.

COROLLARY. *A reflection preserves the set of all lines.*

Let L_1 be any line and L_2 its image under an arbitrary reflection, ρ. Let ρ_1 be the reflection defining L_1. Then, since reflections are involutory, $\rho\rho_1\rho$ leaves just the points on L_2 fixed and reverses orientation. Therefore $\rho\rho_1\rho$ is a reflection and L_2 a line.

4. **Properties of the line.** (A) *Let ρ be a reflection and (P_1, P_2) a pair of points interchanged by ρ (Theorem 2). Then any simple Jordan arc, k, which joins P_1 and P_2 meets the line, L, defined by ρ.*

Since ρ is involutory (Theorem 2), it interchanges k with its image, k'. As a point, Q, traces k from P_1 to P_2, its image, Q', under ρ traces k' from P_2 to P_1. As the arc P_1Q on k increases, Q reaches a first position, Q_1, which is on the image, under ρ, of the arc P_1Q, end points included. Let Q_2 be the image of Q_1. If Q_1 is not on L, the arcs Q_1Q_2 on k and k', respectively, have only their end points in common. Since ρ interchanges these arcs and reverses orientation, it must leave Q_1 and Q_2 fixed. Therefore $Q_1(\equiv Q_2)$ is on L.

LEMMA. *A line, L, is locally connected.*

Under a contrary assumption, let A denote a point at which L is not locally connected; and let c denote an (x, y)-circle about A, so small that every neighborhood of A contains points of L not connected with A on L inside c. In particular, consider the neighborhood N of Lemma 1 in §2, where ϵ is the radius of c. Let k be a simple arc in N, with only its end points on L, joining two points not connected on L inside c. Then, by (A), k and its image, k', under the reflection defining L are distinct and are separated by the points of L inside $(k+k')$. Hence these points of L join the end points of k; but this is contradictory, since, by definition of N, $(k+k')$ is entirely inside c.

(B) *If c is a simple arc with just its end points on a given line, L, then c and its image, γ, under the reflection, ρ, defining L, form a simple closed curve. The points of L inside $(c+\gamma)$ constitute a simple Jordan arc joining the common end points of c and γ.*

From (A), $(c+\gamma)$ is a simple closed curve. Let R be the set of all points which can be connected with c by a simple arc inside $(c+\gamma)$ containing no points of L. Let λ denote the set of points common to L and the boundary of R. The image, R', of R under ρ obviously consists of all the points which can be connected with γ inside $(c+\gamma)$ without meeting L. By (A), R and R' are distinct. Therefore λ separates c from γ and connects their common end points. Also, $(R+\lambda+R')$ contains all points inside $(c+\gamma)$, for otherwise some

point would belong to a finite region bounded solely by points on L, and this region would go into itself under ρ in contradiction with (A). Now $(c+\lambda)$ divides the (x, y)-plane into two regions, the finite region R and the region consisting of R' plus γ plus the exterior of $(c+\gamma)$; and $(c+\lambda)$ is the common boundary of these two regions. Hence, using the above Lemma and §2(A), $(c+\lambda)$ is a simple closed curve, and λ is an arc thereof.

(C) *A line, L, is homeomorphic with the x-axis.*

Let c_1 be a simple arc with just its end points (A_1, B_1) on L, and let Q be a point of c_1. Then c_1 plus its image, γ_1, under the reflection, ρ, defining L cuts from L a Jordan arc, λ_1 (see (B)). Let λ_1 be put in continuous one-to-one correspondence with the segment $-1 \leqq x \leqq 1$ on the x-axis, A_1 corresponding to -1 and B_1 to $+1$.

Proceeding inductively, for $i=1, 2, \cdots$, let c_{i+1} be a simple arc which passes through Q, has only its end points on L, and lies outside $(c_i+\gamma_i)$. Further, let c_{i+1} pass through no point within some positive preassigned distance, d, of the points A_i and B_i. The curve c_{i+1} plus its image, γ_{i+1}, under ρ cuts from L a simple arc, λ_{i+1}, containing λ_i. Let λ_{i+1} be put in continuous one-to-one correspondence with the interval $-(i+1) \leqq x \leqq (i+1)$ on the x-axis in such a way as to preserve the correspondence of λ_i with the interval $-i \leqq x \leqq i$. Let (A_{i+1}, B_{i+1}) be the end points of λ_{i+1} which correspond to $-(i+1)$ and $(i+1)$, respectively. By the last condition imposed on c_{i+1}, neither of the series $(A_1, A_2, \cdots)$ and $(B_1, B_2, \cdots)$ has a cluster point on p. Hence the above inductive process leads to a continuous one-to-one correspondence between the x-axis and a portion, L', of L where L' divides p into two parts. Now the set of all points each inside at least one of the curves $(c_i+\gamma_i)$ $(i=1, 2, \cdots)$ is a neighborhood of L' free from other points of L. Hence L' and $(L-L')$ are distinct. If the set $(L-L')$ is not vacuous, let c' be an arc with just its end points on L, one end point being on L' and one on $(L-L')$. By (B), the end points of c' are joined on L by a simple arc. This contradiction establishes that $L' \equiv L$.

(D) *If two lines, L and L', through any point, A, have any other point, B, of the neighborhood N* (§2, *Lemma* 1) *in common, they coincide.*

Let γ_0 be a curve about A satisfying Theorem 1 and not enclosing B. Suppose γ_0 passes through no common point of L and L'. As the arc AB on L is traced from A, let B' be the first point reached outside γ_0 and on L'. Then B' is joined to γ_0 by two distinct arcs, k and k', on L and L', respectively. Let c be the simple closed curve about A formed of k, k' and an arc on γ_0. Using this for the curve c of Lemmas 2, 3, etc., in §2, we are led to a curve, γ, satisfying Theorem 1. This curve passes through B'. Suppose it

does not. Then A and B' are both inside γ; and, since L and L' meet γ each in just two points, the arcs AB' on L and L' respectively, are both inside γ. Hence c and γ meet only on γ_0. But then* $\gamma=\gamma_0$, which is impossible. Since, therefore, γ passes through the common point B' of L and L', the product of the reflections defining L and L' is the identity (§3, Lemma 1 and Corollary), and $L\equiv L'$ (Theorem 2).

5. **Further properties. The connection with euclidean plane geometry.** The remaining developments prepare for a complete deduction of euclidean plane geometry.

LEMMA 1. *Let* (L_0, L_1) *be two different lines through any point,* A, *and let* (A_j, B_j) $(j=0, 1)$ *be the two points* (§3, *Lemma* 2) *in which* L_j *meets the curve* γ *of Theorem* 1. *Then the points* (A_0, B_0) *separate* A_1 *from* B_1 *on* γ.

Suppose the contrary, and let γ' denote the arc A_0B_0 on γ containing A_1 and B_1. Under the reflection defining L_1, L_0 goes into a line, L_2, which meets γ in the images (A_2, B_2) of (A_0, B_0), and, by §4(A), both A_2 and B_2 must be on the arc A_1B_1 of γ'. Proceeding inductively for $i=1, 2, \cdots$, let L_{2i+1} and L_{2i+2} be the images of L_1 and L_0, respectively, under the reflection, ρ_{2i}, defining L_{2i}. Let (A_{2i+1}, A_{2i+2}) be the images under ρ_{2i} of (A_1, A_0) and (B_{2i+1}, B_{2i+2}) the images of (B_1, B_0). Then the arc A_jB_j on γ' contains A_{j+1} and B_{j+1}. Thus the series $(A_1, A_2, \cdots)$ converges monotonically on γ' to a limit, X. Hence, for i large enough, A_{2i+1} and A_{2i+2} are within an arbitrary given distance of X. By Axiom 3, since ρ_{2i} belongs to T_A, T_A contains a transformation carrying (A_0, A_1) into X. But our transformations are all one-to-one. This contradiction establishes the desired result.

Let (A, P) denote any pair of distinct points on the plane p. The set of all images of P under reflections leaving A fixed will be called a *circle*, with A as *center*.

THEOREM 3. *A circle,* K, *is a simple closed curve.*

(I) Suppose the point P of the above definition is in a neighborhood, N, of A satisfying Lemma 1 in §2, so that K is in a finite region of the (x, y)-plane. Each point, P', on K is image of P under just one reflection leaving A fixed. For suppose there were two such reflections (ρ_1, ρ_2) carrying P into P'. Let ρ be the reflection defining the line AP'. Then $\rho_1\rho\rho_2$ and $\rho_1\rho\rho_1$ are both reflections leaving A and P fixed. Hence they both define the unique line (§4(D)) through A and P. Therefore $\rho_1\rho\rho_2\equiv\rho_1\rho\rho_1$ (Theorem 2), or $\rho_1\equiv\rho_2$.

Using the notation of Lemma 1 above, let γ' be one of the two arcs A_0B_0 on γ. As a point, Q, traces γ' from A_0 to B_0, the line through A and Q adopts

* Since γ is the set of images under T_A of points common to c and γ (Theorem 1, and Lemmas 2, 3 in §2), and T_A preserves γ_0.

the position of every line through A once and only once (Lemma 1), except that the same line is obtained for $Q \equiv A_0$ as for $Q \equiv B_0$. To eliminate the exception, regard A_0 and B_0 as identical, so that Q traces essentially a simple closed curve. Then the image, P', of P under the reflection defining the line AQ adopts every position on K once and only once as Q traces γ'. This affords a one-to-one correspondence between the points on K and those on γ', the point $A_0(\equiv B_0)$ included. It remains only to show the continuity of this correspondence. Let Q be any point on γ' and $(Q_1, Q_2, \cdots)$ a series of points converging to Q. Let P_i $(i=1, 2, \cdots)$ be the point on K corresponding to Q_i and let P^0 be any cluster point of the P's. Then there are transformations leaving A fixed, carrying points arbitrarily near Q into themselves and carrying P into an arbitrary neighborhood of P^0. Hence (Axiom 3) there is a transformation leaving A and Q fixed and carrying P into P^0. This must be the reflection in the line AQ (§4 (D); §3, Lemma 3 and Theorem 2). Therefore P^0 is the point corresponding to Q. This completes the argument for this special case.

(II) Suppose the theorem false. Then, by (I), as some line, L, is traced from A in one sense or the other, a point, P, is reached which is either the last position for which K is a simple closed curve, or the first for which it is not. In the first case, by a proof like that of Lemma 1 in §2, K has a neighborhood consisting of points which remain within distance ϵ of K under all transformations of the set* T_A. Since this neighborhood contains P it follows from (I) that P cannot be the last point for which K is a simple closed curve.

We deal with the second case by showing that if every internal point of the arc AP on L generates a circle which is a simple closed curve, then the circle, K, generated by P is also a simple closed curve. Let it first be noted that no two different lines (L_1, L_2) through A can meet at a point, Q, on K. If they did, then, by §4(D), the arcs AQ on L_1 and L_2 would be distinct and hence form a simple closed curve, c. But then every line determined by A and a point inside c would clearly pass through Q. By reflections in L_1 and L_2 one can show that other lines through A pass through Q; indeed, one can show that all lines through A pass through Q, from which it is easy to deduce a contradiction. Now, for any $\epsilon>0$, there exists a neighborhood, N_ϵ, of P, such that no image of P under the reflection defining a line through a point of N_ϵ is at distance greater than ϵ from P. This may be established by an argument similar to that of Lemma 1 in §2. Consider the correspondence employed in (I) above. We have seen that it is one-to-one even in the present case. We can establish its continuity by an argument like that in (I) applied

* The proof of Corollary 1 below shows that these transformations preserve K.

to images of P in N_ϵ, and, similarly, in a neighborhood of any point on K. Hence K is a simple closed curve.

COROLLARY 1. *A circle with center at A is preserved by all the transformations T_A (Axiom 3).*

Let τ be a transformation of the set T_A. Let P'' be the image under τ of any point P' on K, and let ρ be the reflection which leaves A fixed and interchanges (P, P'). Then, if ρ' is the reflection defining the line AP', $\rho\tau$ and $\rho\rho'\tau$ both carry P into P'' and leave A fixed. The one which reverses orientation is a reflection, for it preserves the curve γ of Theorem 1 and therefore leaves two of its points fixed. Hence P'' is on K.

COROLLARY 2. *A circle is met in just two points by a line through its center.* (See §3, Lemma 2.)

COROLLARY 3. *Any two points (P_1, P_2) on a circle are interchanged by some reflection leaving the center fixed.*

Let ρ_j be the reflection leaving the center, A, fixed and carrying P into P_j $(j=1, 2)$. If ρ define the line through A and P, the product $\rho_1\rho\rho_2$ reverses orientation, carries P_1 into P_2 and preserves the circle. It is therefore a reflection, for it leaves two points on the circle fixed. Hence (Theorem 2) it interchanges P_1 and P_2.

THEOREM 4. *One and only one line passes through any two distinct points of the plane p.*

If two lines through a point, A, have a second point, P, in common, their defining reflections preserve the circle through P with center at A and leave P fixed. The product of these reflections is therefore the identity and the lines coincide (§3).

COROLLARY 1. *For any two points, A and B, on the plane p, there exists a reflection interchanging them.*

Let P be a common point of the two circles, centers at A and B, respectively, passing through B and A, respectively. Let ρ_1 be the reflection interchanging A and P and leaving B fixed (Theorem 3, Corollary 3) and ρ_2 the reflection interchanging B and P and leaving A fixed. Then $\rho_1\rho_2\rho_1$ reverses orientation, leaves P fixed, and carries A into B. Since ρ_1 and ρ_2 are involutory (Theorem 2), $\rho_1\rho_2\rho_1$ is also involutory and therefore leaves more than one point fixed (see proof of §4(A)). Hence it is the required reflection.

COROLLARY 2. *Let L and L' be any two lines. Let A be any point on L and B any point on L'. Then some transformation in the group T carries L into L' in such a way that A goes into B.*

The reflection, ρ_1, which interchanges A and B (Corollary 1) carries L into a line, L'', through B (Theorem 2, Corollary). Some reflection, ρ_2, by Theorem 3, Corollary 3, leaves B fixed and interchanges any two points in which L' and L'', respectively, meet a circle, center at B. The product $\rho_1\rho_2$ satisfies the requirements of the present corollary.

There remain no difficulties in defining angles, distances, congruences, and proceeding with other geometrical developments, or else establishing the axioms in Chapter 1 of Hilbert's *Grundlagen der Geometrie*.

Two geometries rest on the above foundation: the euclidean if one assume that through a given point not on a given line, L, there is but one line which fails to meet L; the Bolyai-Lobatchewsky if one assume that there are two lines through the point separating the intersecting from the non-intersecting lines.

Lehigh University,
Bethlehem, Pa.

PROOF OF THE FUNDAMENTAL THEOREMS ON SECOND-ORDER CROSS PARTIAL DERIVATIVES*

BY

A. E. CURRIER

1. Introduction. In the present article we prove the following theorem:

Let $f(x, y)$ be defined on an open region R, and let the first partial derivatives f_x and f_y exist on R. Let A be a point set on which the four second-order partial derivatives

$$f_{xx}, f_{xy}, f_{yx}, f_{yy} \tag{1.1}$$

exist almost everywhere. Then

$$f_{xy} = f_{yx}$$

almost everywhere on A.

At the conclusion of this paper we state a number of interesting problems which are closely connected with the theorems which we prove. We wish to thank Dr. S. Saks for many helpful suggestions.

2. Measurability of second partial derivatives. We require the following lemma.

LEMMA 1. *Let $u(x, y)$ be a function of Baire, and let A_n be the point set on which the following inequality is satisfied:*

$$\frac{u(x+h, y) - u(x, y)}{h} < \alpha, \quad 0 < h < \frac{1}{n}, \quad (x, y) < A_n. \tag{2.1}$$

Then A_n is measurable.

We readily see that the function

$$\frac{u(x+h, y) - u(x, y)}{h}$$

is a function of Baire in the space of the three variables (x, y, h). Let Δ be the portion of the space (x, y, h) for which $0 < h < 1/n$. Let $\mathfrak{A}_n$ be the portion of Δ for which the inequality (2.1) is satisfied. We see that the complement of the

* Presented to the Society, December 27, 1932; received by the editors, May 20, 1932, and, in revised form, June 8, 1932.

set A_n of Lemma 1 is obtained by projecting $\Delta - \mathfrak{A}_n$ onto the space (x, y). Thus A_n is a complementary analytic set, and hence measurable.*

From Lemma 1 the reader will readily see that the following lemma is true.

LEMMA 2. *Let $f(x, y)$ be a function of Baire, defined on an open region R. Let the first partial derivative f_x exist on R. Then the second partial derivatives f_{xx} and f_{xy} are measurable functions on their respective domains of definition.*

3. **Cross sections of a closed point set.** We make the following definition.

DEFINITION. *If P is a closed point set in the plane (x, y) then $P_x(t)$ and $P_y(t)$ will denote the cross sections of P with the lines $x=t$ and $y=t$ respectively.*

We shall require the following lemma.

LEMMA 3. *If P is closed† then almost every point (x_0, y_0) of P is a density point in the linear sense of the cross sections $P_x(x_0)$ and $P_y(y_0)$, and in the superficial sense a density point of P.*

We denote by $\phi(x, y)$ the *characteristic function*‡ of P and set

$$\psi(x, y) = \int_0^y \phi(x, y)dy.$$

By a well known theorem the partial derivative ψ_y exists and equals $\phi(x, y)$ almost everywhere in the plane. That is, $\psi_y(x, y) = 1$ almost everywhere on P. This shows that almost every point of P is a density point in the linear sense of the corresponding cross section $P_x(x)$. A similar proof applies to the cross sections $P_y(y)$. It is well known that the superficial density points of P lie almost everywhere on P, hence Lemma 3 is correct.

4. **The approximate middle derivative.** The *middle difference quotient* $\Delta f/\lambda^2$ of a function $f(x, y)$ at the point (x_0, y_0) is defined as follows:

$$(4.1) \quad \frac{\Delta f}{\lambda^2} = \frac{f(x_0 + \lambda, y_0 + \lambda) - f(x_0, y_0 + \lambda) - f(x_0 + \lambda, y_0) + f(x_0, y_0)}{\lambda^2}.$$

The *approximate middle derivative* of $f(x, y)$ at (x_0, y_0) (when it exists) will be defined as the following *approximate limit*§:

---

* N. Lusin, *Sur les ensembles analytiques*, Fundamenta Mathematicae, vol. 10 (1927), pp. 1–95, especially pp. 25–26. Also Lusin et Sierpinski, *Sur quelques propriétés des ensembles* (A), Bulletin de l'Académie de Cracovie, 1918, p. 44.

† The lemma holds if P is merely measurable.

‡ The characteristic function of a point set equals one on points of the set and equals zero elsewhere.

§ Cf. Lebesgue, *Leçons sur l'Intégration*, Paris, 1928, pp. 240–241, for a definition of approximate limits.

$$\text{approx-}D_m f = \text{approx-}\lim_{\lambda \to 0} \frac{\Delta f}{\lambda^2} \cdot \tag{4.2}$$

5. The fundamental lemma. We now state the fundamental lemma as follows.

FUNDAMENTAL LEMMA. *Let $f(x, y)$ be a function of Baire defined on an open region R, and let the first partial derivative $f_x(x, y)$ exist on R. Let A be the subset of R on which the partial derivatives*

$$f_{xx}, f_{xy}$$

*exist and take on finite values. Then**

$$\text{approx-}D_m f = f_{xy} \tag{5.1}$$

almost everywhere on A.

The functions f_{xx} and f_{xy} are measurable by Lemma 2, hence the set A is measurable. Let A_n be the part of A for which the following inequalities are satisfied:

$$\left| \frac{f_x(x + h, y) - f_x(x, y)}{h} \right| < n, \tag{5.2}$$

$$\left| \frac{f_x(x, y + k) - f_x(x, y)}{k} \right| < n, \tag{5.3}$$

for $(x, y) < A_n$ and $0 < |h|, |k| < 1/n$.

We readily see that the sets A_n cover the set A. Hence in order to prove the Fundamental Lemma it is sufficient to show that (5.1) holds almost everywhere on the set A_n.

By Lemma 1 the set A_n is seen to be measurable, and hence by a well known theorem there exists a sequence $\{P_k\}$ of closed parts of A_n which cover A_n almost everywhere. Hence in order to prove the Fundamental Lemma it is sufficient to show that (5.1) holds almost everywhere on each closed part of A_n.

Let P be a closed part of A_n. From (5.3) it follows that the function f_{xy} is bounded on P. Since f_{xy} is measurable and bounded on P it is summable on P and by a well known theorem for almost every point (x_0, y_0) of P

$$\lim_{\lambda \to 0} \frac{1}{\lambda^2} \int_{P\delta} f_{xy} dx dy = f_{xy}(x_0, y_0), \tag{5.4}$$

---

* Equation (5.1) implies the existence of approx-$D_m f$, that is, approx-$D_m f$ exists and equals f_{xy} almost everywhere on A.

where δ is the square with one corner at the point (x_0, y_0) and the opposite corner at the point $(x_0+\lambda, y_0+\lambda)$. Moreover by Lemma 3 almost every point (x_0, y_0) of P is a density point in the linear sense of the cross sections of P and in the superficial sense a density point of P. Hence in order to prove that (5.1) holds almost everywhere on P it is sufficient to show that (5.1) holds for the point (x_0, y_0) of P, where (x_0, y_0) is a density point of P in the linear and superficial senses and is also a point for which (5.4) holds. Thus we see that in order to prove the Fundamental Lemma it is sufficient to prove the following auxiliary lemma.

AUXILIARY LEMMA. *Let the hypotheses of the Fundamental Lemma be satisfied. Let P be a closed part of A such that the inequalities* (5.2) *and* (5.3) *are satisfied for (x, y) on P.*

Let (x_0, y_0) be a density point in the linear sense of the cross sections of P and in the superficial sense a density point of P. Moreover let (x_0, y_0) be a point of P at which equation (5.4) *holds.*

Then approx-$D_m f$ *exists at (x_0, y_0) and equals $f_{xy}(x_0, y_0)$.*

6. **Proof of the Auxiliary Lemma.** Let e be the set of values of λ corresponding to which the points $(x_0, y_0+\lambda)$ lie in P. We see that the point $\lambda=0$ is a density point of e, since (x_0, y_0) is a density point in the linear sense of the cross section $P_x(x_0)$ of P. For λ in e and constant we see by (5.2) that $f_x(x, y_0+\lambda)$ is uniformly bounded for $|x-x_0|<1/n$. Hence the middle increment Δf at (x_0, y_0) can be expressed by means of an integral as follows:

$$\Delta f = \int_{x_0}^{x_0+\lambda} [f_x(x, y_0+\lambda) - f_x(x, y_0)]dx. \tag{6.1}$$

As before, let δ be the closed square with one corner at (x_0, y_0) and the opposite corner at $(x_0+\lambda, y_0+\lambda)$. Let e_1 be the projection of $P\delta$ on the x-axis, and let e_2 be the complement of e_1 with respect to the closed interval $x_0 \leqq x \leqq x_0+\lambda$ (or $x_0+\lambda \leqq x \leqq x_0$ if $\lambda<0$).

Let O denote the ordinate which passes through the point $(x, 0)$. Then the product set OP is the cross section $P_x(x)$ of P. The set $OP\delta$ is closed and non-empty for x in e_1. The complementary set $O\delta - OP\delta$ is open on $O\delta$ and consists of at most a denumerable infinity of open linear intervals on the ordinate O. Let (x, α) and (x, β) be the end points of a general one of these intervals. In the difference quotient

$$\frac{f_x(x, \beta) - f_x(x, \alpha)}{\beta - \alpha}$$

at least one of the points (x, α), (x, β) is a point of P. Moreover $|\beta-\alpha| < |\lambda|$.

Hence by (5.3) we have

$$(6.2)\qquad |f_x(x,\beta)-f_x(x,\alpha)|<n|\beta-\alpha|,\ |\lambda|<\frac{1}{n}\cdot$$

Hence for x in e_1 and $|\lambda|<1/n$ the difference $f_x(x, y_0+\lambda)-f_x(x, y_0)$ can be expressed as follows:*

$$(6.3)\ f_x(x,y_0+\lambda)-f_x(x,y_0)=\int_{OP\delta} f_{xy}(x,y)dy+\sum_{\alpha,\beta}[f_x(x,\beta)-f_x(x,\alpha)].$$

Comparing (6.1) and (6.3) we see that for λ in e and $|\lambda|<1/n$ the middle increment Δf can be expressed in the form

$$(6.4)\qquad \begin{aligned}\Delta f=\int_{e_1}dx\int_{OP\delta} f_{xy}(x,y)dy&+\int_{e_1}\sum_{\alpha,\beta}[f_x(x,\beta)-f_x(x,\alpha)]dx\\ &+\int_{e_2}[f_x(x,y_0+\lambda)-f_x(x,y_0)]dx.\end{aligned}$$

As remarked above the point $\lambda=0$ is a density point of e. Hence in order to prove the Auxiliary Lemma it is sufficient to show that

$$(6.5)\qquad \lim_{\lambda\to_e 0}\frac{\Delta f}{\lambda^2}=f_{xy}(x_0,y_0)$$

where the notation on the left indicates that λ is to approach zero through values which lie in e.

The first integral on the right of (6.4) is merely the double integral of $f_{xy}(x, y)$ taken over $P\delta$. Because of (5.4) we see that in order to prove (6.5) it is sufficient to show that the following equations are true:

$$(6.6)\qquad \lim_{\lambda\to_e 0}\frac{1}{\lambda^2}\int_{e_1}\sum_{\alpha,\beta}[f_x(x,\beta)-f_x(x,\alpha)]dx=0,$$

$$(6.7)\qquad \lim_{\lambda\to_e 0}\frac{1}{\lambda^2}\int_{e_2}[f_x(x,y_0+\lambda)-f_x(x,y_0)]dx=0.$$

Because of (5.3) we see that the absolute value of the sum in the integrand (6.6) is bounded by

$$n\sum_{\alpha,\beta}|\beta-\alpha|=n\mu(O\delta-OP\delta)$$

* Cf. Lebesgue, loc. cit., pp. 210–211. The series in (6.3) is readily seen to be absolutely convergent by (6.2).

where μ denotes linear Lebesgue measure on the ordinate O. Hence the absolute value of the expression (6.6) is bounded by the following expression:

$$(6.8)\quad \frac{1}{\lambda^2}\int_{e_1} n\mu(O\delta - OP\delta)dx \leqq \frac{1}{\lambda^2}\int_{x_0}^{x_0+\lambda} n\mu(O\delta - OP\delta)dx = n\frac{m(\delta - P\delta)}{\lambda^2},$$

where m denotes superficial Lebesgue measure. Since (x_0, y_0) is a superficial density point of P the expression on the right of (6.8) converges to zero with λ. Hence (6.6) is correct.

To prove (6.7) we rewrite the integrand as follows:

$$(6.9)\quad \begin{aligned} f_x(x, y_0 + \lambda) - f_x(x, y_0) &= f_x(x, y_0 + \lambda) - f_x(x_0, y_0 + \lambda) \\ &\quad + f_x(x_0, y_0 + \lambda) - f_x(x_0, y_0) + f_x(x_0, y_0) - f_x(x, y_0). \end{aligned}$$

Thus we see that in order to prove (6.7) it is sufficient to show that the following three equations hold:

$$(6.10)\quad \lim_{\lambda \to_e 0} \frac{1}{\lambda^2}\int_{e_2} [f_x(x, y_0 + \lambda) - f_x(x_0, y_0 + \lambda)]dx = 0;$$

$$(6.11)\quad \lim_{\lambda \to_e 0} \frac{1}{\lambda^2}\int_{e_2} [f_x(x_0, y_0 + \lambda) - f_x(x_0, y_0)]dx = 0;$$

$$(6.12)\quad \lim_{\lambda \to_e 0} \frac{1}{\lambda^2}\int_{e_2} [f_x(x_0, y_0) - f_x(x, y_0)]dx = 0.$$

For λ in e the point $(x_0, y_0+\lambda)$ lies in P and for λ in e and $|\lambda| < 1/n$ we see from (5.2) that the absolute value of the expression (6.10) is bounded by

$$(6.13)\quad \frac{1}{\lambda^2}|\lambda| n\nu e_2 = n\left|\frac{\nu e_2}{\lambda}\right|,$$

where ν denotes linear Lebesgue measure in the x-space. From the fact that (x_0, y_0) is a density point in the linear sense of the cross section $P_y(y_0)$ of P we readily see that the expression on the right of (6.13) converges to zero as λ approaches zero. Hence (6.10) is correct. In a similar way we prove that (6.11) and (6.12) are correct. Hence the Auxiliary Lemma and the Fundamental Lemma of §5 are correct.

7. **The first theorem on second-order partial derivatives.** We now prove the following theorem.

THEOREM 1. *Let $f(x, y)$ be a function of Baire defined on an open region R, and let the first partial derivative $f_x(x, y)$ exist on R. Let A be a point set on which the partial derivatives f_{xx} and f_{xy} exist almost everywhere. Then the approximate middle derivative* approx-$D_m f$ *exists almost everywhere on A and*

$$\text{approx-}D_m f = f_{xy} \tag{7.1}$$

almost everwhere on A.

Let A' be the part of A on which f_{xx} and f_{xy} exist and take on finite values. The set $A-A'$ has superficial measure zero,* hence in order to prove Theorem 1 it is sufficient to show that (7.1) holds almost everywhere on A'. Let A'' be the set on which f_{xx} and f_{xy} exist and take on finite values. The set A' is part of A''. By the Fundamental Lemma approx-$D_m f$ exists and equals f_{xy} almost everywhere on A'', hence almost everywhere on A'. Thus Theorem 1 is correct.

8. **Equality of the cross partial derivatives.** We now prove the following:

THEOREM 2. *Let $f(x, y)$ be defined on an open region R, and let the first partial derivatives $f_x(x, y)$ and $f_y(x, y)$ exist on R. Let A be a point set on which the four second-order partial derivatives $f_{xx}, f_{xy}, f_{yx}, f_{yy}$ exist almost everywhere. Then $f_{xy} = f_{yx}$ almost everywhere on A.*

Since f_x and f_y exist, the function $f(x, y)$ is continuous in x alone and continuous in y alone, and is thus a function of Baire. By Theorem 1 approx-$D_m f$ $=f_{xy}$ almost everywhere on A. By reasoning similar to the proof of Theorem 1 we see that approx-$D_m f=f_{yx}$ almost everywhere on A. That is

$$f_{xy} = \text{approx-}D_m f = f_{yx}$$

almost everywhere on A. Thus Theorem 2 is seen to be true.

9. **Generalizations of Theorems 1 and 2.** In Theorems 1 and 2 it is not necessary to assume that the first derivatives exist. We make the following definition:

DEFINITION 1. *Let $f(x, y)$ be an arbitrary function, and let*

$$\bar{f}_x^{+}(x, y),\ \underline{f}_x^{+}(x, y),\ \bar{f}_x^{-}(x, y),\ \underline{f}_x^{-}(x, y) \tag{9.1}$$

be the four principal first partial derivatives† of $f(x, y)$ with respect to x. The second partial derivative f_{xx} is said to exist at (x_0, y_0) if the first derivative $f_x(x_0, y_0)$ exists (that is, if the four functions (9.1) *have the same value at (x_0, y_0)) and if the four functions* (9.1) *are partially differentiable with respect to x at (x_0, y_0) and have the same value for their partial derivative with respect to x at (x_0, y_0).*

* The set $A-A'$ consists of the part of A on which f_{xx} or f_{xy} does not exist (this part is of measure zero) plus the part of A on which either f_{xx} or f_{xy} is infinite. The latter part also has measure zero. The function f_{xx} is measurable, and the linear measure of each cross section of the part of A on which $f_{xx}=\pm\infty$ is zero, as proved for example in Hobson, *Theory of Functions of a Real Variable*, Cambridge, 1927, vol. I, p. 397, Theorem 2. Similarly for infinite values of f_{xy}.

† Cf. Carathéodory, *Vorlesungen über reelle Funktionen*, Berlin, 1927, p. 641.

We make similar definitions for the existence of the remaining second partial derivatives. We now state certain generalizations of Theorems 1 and 2.

THEOREM 3. *Let $f(x, y)$ be defined on an open region R, and be continuous in x alone and in y alone. Let $D_x f$ be one of the principal first partial derivatives* (9.1). *Let A be a point set on which the partial derivatives*

$$\frac{\partial}{\partial x} D_x f, \quad \frac{\partial}{\partial y} D_x f$$

exist almost everywhere. Then

$$\text{approx-}D_m f = \frac{\partial}{\partial y} D_x f \tag{9.2}$$

almost everywhere on A.

It is well known that under the hypotheses of Theorem 3 the function $D_x f$ is a function of Baire. The proof of Theorem 3 now follows in the same way as the proof of Theorem 1, after a suitable restatement and reproof of the Fundamental Lemma. No new methods are required.

THEOREM 4. *Let $f(x, y)$ be defined on an open region R, and let $f(x, y)$ be continuous in x alone and in y alone. Let A be a point set on which the second partial derivatives*

$$f_{xx}, \ f_{xy}, \ f_{yx}, \ f_{yy}$$

exist in the generalized sense almost everywhere. Then

$$f_{xy} = f_{yx}$$

almost everywhere on A.

The proof of this theorem follows readily, making use of Theorem 3.

10. **Problems.** Certain problems present themselves at once. We state a few of them here.

PROBLEM 1. *Let $f(x, y)$ be a function of Baire. Are the principal first partial derivative functions* (9. 1) *also functions of Baire**?

If the answer to Problem 1 is in the affirmative the following theorem follows at once by the methods of this paper.

---

* For a function $f(x)$ of one variable the answer is in the affirmative; cf. Sierpinski, *Sur les fonctions dérivées des fonctions discontinues*, Fundamenta Mathematicae, vol. 3 (1922), pp. 123–127. In case $f(x,y)$ is continuous in x the answer is in the affirmative.

Let $f(x, y)$ be a function of Baire. Let A be a point set on which the second partial derivatives f_{xx}, f_{xy}, f_{yx}, f_{yy} exist in the generalized sense almost everywhere. Then $f_{xy}=f_{yx}$ almost everywhere on A.

PROBLEM 2. *Let $f(x, y)$ together with its first partial derivatives f_x and f_y be continuous on an open region R. Let A be a point set on which f_{xx} and f_{xy} exist almost everywhere. Then we know that* approx-$D_m f$ *exists and equals f_{xy} almost everywhere on A. Does the actual middle derivative $D_m f$ exist almost everywhere on A?*

PROBLEM 3. *Is Theorem* 1 *true if $f(x, y)$ is merely measurable?*

It is probable that an example can be constructed which will show that the answer to this problem is in general negative.

A large number of related problems can be readily thought of, problems which have to do with the reversal of the order of integration in iterated integrals, and problems connected with two-dimensional totalization. It is of course quite obvious that theorems similar to Theorems 1, 2, 3, and 4 can be stated concerning the third partial derivatives f_{xyz} etc. of a function $f(x, y, z)$ of three variables, and corresponding general theorems for the partial derivatives of functions of n variables.

HARVARD UNIVERSITY,
CAMBRIDGE, MASS.

THE CANCELLATION LAW IN THE THEORY OF CONGRUENCES TO A DOUBLE MODULUS*

BY

MORGAN WARD

1. Let m be an integer greater than unity and $f(x)$ a fixed polynomial with integral coefficients.† If the leading coefficient of $f(x)$ is prime to m, then the quotient and remainder obtained on dividing any other polynomial by $f(x)$ have integral coefficients modulo m. Hence, as is well known, all polynomials may be separated into a finite number of residue classes $\mathfrak{A}, \mathfrak{B}, \cdots$, $\mathfrak{U}, \cdots$ which form a commutative ring‡ with respect to the operations of addition and multiplication (modulis $m, f(x)$). I propose here to determine what inferences can be drawn concerning the ring elements $\mathfrak{U}$ and $\mathfrak{B}$ from the ring equality $\mathfrak{A}\mathfrak{U} = \mathfrak{A}\mathfrak{B}$ when $\mathfrak{A} \neq 0$. Since $\mathfrak{A}\mathfrak{U} = \mathfrak{A}\mathfrak{B}$ is equivalent to $\mathfrak{A}(\mathfrak{U} - \mathfrak{B}) = 0$, we may assume that $\mathfrak{B} = 0$. Stated in terms of congruences our problem is then equivalent to the following one:

Suppose that $f(x) = c_0x^k + c_1x^{k-1} + \cdots + c_k$ is a fixed polynomial with integral coefficients $c_0, \cdots, c_k$ and that m is an integer prime to c_0. Let $A(x)$ be a given polynomial such that

$$A(x) \not\equiv 0 \qquad (\text{modd } m, f(x)).$$

To determine all polynomials $U(x)$ such that

$$A(x)U(x) \equiv 0 \qquad (\text{modd } m, f(x)). \tag{1.1}$$

The problem is essentially a generalization of the problem of solving

$$au \equiv 0 \qquad (\text{mod } m)$$

for given integers a and m. Nevertheless it does not seem to have been considered heretofore save in very special cases.

I shall first of all show that it is sufficient to consider the case when m is a power of a prime p, say $m = p^N$, and when $f(x)$ is congruent modulo p to a power of an irreducible polynomial $\phi(x)$ (mod p);

$$f(x) = B(x) \equiv \{\phi(x)\}^\beta \qquad (\text{mod } p).$$

* Presented to the Society, August 31, 1932; received by the editors May 24, 1932.

† We shall restrict the term polynomial in what follows to mean a polynomial with integral coefficients.

‡ van der Waerden, *Moderne Algebra*, Berlin, 1930, vol. I, p. 37; Haupt, *Einführung in die Algebra*, Leipzig, 1929, vol. I, chapter V.

This reduction corresponds to the fact that the ring associated with the moduli m and $f(x)$ is the direct sum of rings of the type associated with the moduli p^N and $B(x)$.

In this simpler case I shall show that there exists a positive integer λ and a set (S) of λ polynomials

$$A_0(x), A_1(x), \cdots, A_{\lambda-1}(x)$$

where λ and (S) depend only upon $A(x)$ and $B(x)$ and are independent of N, such that

$$A(x)U(x) \equiv 0 \qquad (\text{modd } p^N, B(x))$$

when and only when

$$U(x) \begin{array}{l} = p^{N-\lambda}(Q_0(x)A_0(x) + pQ_1(x)A_1(x) + \cdots + p^{\lambda-1}Q_{\lambda-1}(x)A_{\lambda-1}(x)) \text{ if } N = \lambda \\ = Q_{\lambda-N}(x)A_{\lambda-N}(x) + pQ_{\lambda-N+1}(x)A_{\lambda-N+1}(x) + \cdots + p^{N-1}Q_{\lambda-1}(x)A_{\lambda-1}(x) \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if } N \leqq \lambda, \end{array}$$

the polynomials $Q(x)$ being completely arbitrary, save for a restriction upon their degrees which we shall give later.

In the ring associated with the double modulus p^N, $B(x)$, our results are equivalent to the theorem that the ideal to which every element $\mathfrak{U}$ of the ring belongs which satisfies the relation

$$\mathfrak{A}\mathfrak{U} = 0$$

has a basis of the form

$$p^{N-\lambda}\mathfrak{A}_0, p^{N-\lambda+1}\mathfrak{A}_1, \cdots, p^{N-1}\mathfrak{A}_{\lambda-1} \quad \text{if} \quad N \geqq \lambda$$

or of the form

$$\mathfrak{A}_{\lambda-N}, p\mathfrak{A}_{\lambda-N+1}, \cdots, p^{N-1}\mathfrak{A}_{\lambda-1} \quad \text{if} \quad N \leqq \lambda,$$

where λ and $\mathfrak{A}_0, \cdots, \mathfrak{A}_{\lambda-1}$ depend only upon $\mathfrak{A}$, p and $B(x)$ and are independent of N.

2. Suppose that

$$m = p_1^{n_1} \cdots p_r^{n_r}$$

is the decomposition of m into its prime factors. Then it is readily seen that a necessary and sufficient condition that the congruence (1.1) hold is that the r congruences

$$A(x)U(x) \equiv 0 \qquad (\text{modd } p_i^{n_i}, f(x)),\ i = 1, \cdots, r, \tag{2.1}$$

hold. Furthermore, if we know the general solution $U^{(i)}(x)$ of each of the congruences (2.1), the general solution of the congruence (1.1) can be written

down immediately by means of the Chinese remainder theorem.* Hence it is sufficient to discuss the case when $m=p^N$, p a prime.

Let

$$f(x) \equiv c_0\{\phi_1(x)\}^{\beta_1} \cdots \{\phi_s(x)\}^{\beta_s} \qquad (\text{mod } p),\ (c_0, p) = 1,$$

be the decomposition of $f(x)$ into primary irreducible polynomials modulo p. Then by Schönemann's second theorem† there exists a decomposition of $f(x)$ modulo p^N of the type

$$f(x) \equiv c_0' B_1(x) \cdots B_s(x) \qquad (\text{mod } p^N), (c_0', p) = 1,$$

where the polynomials $B_i(x)$ are primary, and

$$B_i(x) \equiv \{\phi_i(x)\}^{\beta_i} \qquad (\text{mod } p), i = 1, \cdots, s. \tag{2.2}$$

Since $\text{Res}\{B_i(x), B_j(x)\}$ is prime to p if $i \neq j$, it easily follows that (1.1) holds with $m=p^N$ when and only when the s congruences

$$A(x)U(x) \equiv 0 \qquad (\text{modd } p^N, B_i(x)), i = 1, \cdots, s,$$

hold. If the solutions of these congruences are known, then the solution of the congruence (1.1) may be written down by the procedure of the Chinese remainder theorem.

It is sufficient then to study the congruence

$$A(x)U(x) \equiv 0 \qquad (\text{modd } p^N, B(x)) \tag{2.3}$$

where

$$A(x) = a_0x^n + a_1x^{n-1} + \cdots + a_n,\ B(x) = x^m + b_1x^{m-1} + \cdots + b_m$$

are given polynomials, p is a prime number, N a positive integer, and $U(x)$ is to be determined. Furthermore

$$B(x) \equiv \{\phi(x)\}^{\beta} \qquad (\text{mod } p) \tag{2.4}$$

where $\phi(x)$ is a primary irreducible polynomial modulo p and β a positive integer. We shall not need to use this last fact in what immediately follows. Finally, we lose no generality by requiring that $U(x)$ be of lesser degree than $B(x)$.

3. The first problem is to determine for a given N the highest power of p which divides every $U(x)$ satisfying (2.3). We shall show that there exists an integer λ depending only upon $A(x)$ and $B(x)$ such that if $N>\lambda$, every solution of (2.3) is divisible by $p^{N-\lambda}$, while if $N \leqq \lambda$, there exist solutions of

* Dickson, *Introduction to the Theory of Numbers*, Chicago, 1929, p. 10.

† For an account of Schönemann's theorems, see Fricke, *Algebra*, Braunschweig, 1928, vol. II, chapter 2.

(2.3) which are not divisible by p. If $N>\lambda$, we may therefore write $U(x)=p^{N-\lambda}W(x)$ and thus reduce (2.3) to a congruence of the same form with $N=\lambda$. We shall see in the next section that the discussion of (2.3) when $N\leqq\lambda$ presents no difficulties whatsoever.

Let

$$E = E\begin{pmatrix} a_0, a_1, \cdots, a_n \\ 1, b_1, \cdots, b_m \end{pmatrix}$$

denote the $(m+n)$-rowed Sylvester eliminant of the polynomials $A(x)$ and $B(x)$, and let

$$\mathcal{E} = (e_{ij}) \qquad (i, j = 1, \cdots, m+n)$$

denote the transpose of the matrix of the determinant E.

Suppose that

$$E = p^L E' \quad \text{where} \quad L \geqq 0, \quad (p, E') = 1.$$

The congruence (2.3) is equivalent to an identity in x of the form

$$A(x)U(x) + B(x)V(x) = p^N W(x)$$

where the polynomial $V(x)$ is at most of degree $n-1$ and the polynomial $W(x)$ at most of degree $m+n-1$. If we denote the $m+n$ unknown coefficients of $U(x)$ and $V(x)$ in order by $z_1, z_2, \cdots, z_{m+n}$ and the coefficients of $W(x)$ by $w_1, w_2, \cdots, w_{m+n}$, then this identity is easily seen to be equivalent to the system of $m+n$ linear equations

$$(3.1) \qquad \sum_{j=1}^{t} e_{ij} z_j = p^N w_i \qquad (i = 1, \cdots, t)$$

where for brevity we have written t for $m+n$. The determinant of this system is E; hence

$$E z_j = p^N \sum_{i=1}^{t} \bar{e}_{ji} w_i \qquad (j = 1, \cdots, t),$$

$\bar{e}_{ji}$ denoting the co-factor of e_{ij} in E. Suppose that p^D is the highest power of p dividing all of the first minors $\pm\bar{e}_{ji}$ of E. Then on writing $p^L E'$ for E, we see that

$$E' z_j = p^{N+D-L} \sum_{i=1}^{t} e'_{ji} w_i \qquad (j = 1, \cdots, t)$$

where $p^D e'_{ji} = \bar{e}_{ji}$. At least one of the numbers e'_{ji} is not divisible by p; suppose that it is e'_{kl}. Then on taking w_l equal to 1 and the remaining w equal to 0,

we obtain a solution of (3.1) such that every z is divisible by p^{N+D-L} and at least one z, namely z_k, is not divisible by any higher power of p. It follows that the highest power of p dividing *all* solutions of (3.1) is p^{N+D-L}.

The integer p^{L-D} is simply the first elementary divisor of the matrix $\mathcal{E}$ corresponding to the prime factor p. Writing λ for $L-D$, we have the following result:

The least value of N such that every solution $U(x)$ of (2.3) of degree less than $B(x)$ should be divisible by p^T is $T+\lambda$, where p^λ is the first elementary divisor corresponding to the prime p of the matrix of the eliminant of $A(x)$ and $B(x)$.

Consequently, if $N\leqq\lambda$, there exist solutions of (2.3) which are not divisible by p, while if $N>\lambda$, every solution is divisible by $p^{N-\lambda}$. Since $\lambda=0$ only when $L=D=0$, we must have $U(x)\equiv 0 \pmod{p^N}$ if the resultant of $A(x)$ and $B(x)$ is prime to p. In the ring associated with p^N and $B(x)$, the corresponding case is when $\mathfrak{A}\mathfrak{U}=0$ and $\mathfrak{A}$ is a unit of the ring.

4. We can now complete the discussion of the congruence (2.3). If $N>\lambda$, set $U(x)=p^{N-\lambda}W(x)$ thus obtaining the congruence for $W(x)$

$$A(x)W(x) \equiv 0 \qquad (\text{modd } p^\lambda, B(x)). \tag{4.1}$$

Among the polynomials $W(x)$ which satisfy (4.1) are some not divisible by p. Let $T(x)$ be such a one of lowest possible degree. Then the leading coefficient of $T(x)$ must be prime to p; for if not, by Schönemann's first theorem* there would exist a polynomial of the form $c+pQ(x)$ where c is prime to p such that $T(x)(c+pQ(x))$ would be congruent modulo p^λ to a polynomial $T'(x)$ of lower degree than $T(x)$. Then, since Res $\{c+pQ(x), B(x)\}$ is prime to p, we would have $A(x)T'(x)\equiv 0$ (modd p^λ, $B(x)$) contradicting our assumption about the degree of $T(x)$. On multiplying $T(x)$ by a constant prime to p, we obtain a polynomial $A_0(x)$ with leading coefficient unity and of minimal degree satisfying (4.1). *This polynomial is unique modulo p^λ*; for the difference of two such polynomials would be of lesser degree than either. Moreover if $W(x)$ is any solution of (4.1), the quotient and remainder obtained on dividing $W(x)$ by $A_0(x)$ have integral coefficients and the remainder being of lower degree than $A_0(x)$ must be divisible by p. Hence

$$W(x) = Q_0(x)A_0(x) + pW_1(x)$$

where $W_1(x)$ is of lesser degree than $A_0(x)$. On substituting this expression in (4.1), we obtain a congruence of the same form for $W_1(x)$:

$$A(x)W_1(x) \equiv 0 \qquad (\text{modd } p^{\lambda-1}, B(x)).$$

* Fricke, loc. cit., p. 59.

We now repeat the previous argument. Every solution of this congruence must be of the form

$$W_1(x) = Q_1(x)A_1(x) + pW_2(x)$$

where $A_1(x)$ is a solution of minimal degree in x with leading coefficient unity uniquely determined modulo $p^{\lambda-1}$, while $W_2(x)$ is of lesser degree than $A_1(x)$.

We find on continuing in this manner that the general solution of (4.1) is of the form

$$W(x) = Q_0(x)A_0(x) + pQ_1(x)A_1(x) + \cdots + p^{\lambda-1}Q_{\lambda-1}(x)A_{\lambda-1}(x)$$

where the polynomial $A_i(x)$ is uniquely determined modulo $p^{\lambda-i}$.

We shall show in the next section that two consecutive polynomials $A_r(x)$ and $A_{r+1}(x)$ are equal only when all the polynomials $A_r(x)$, $A_{r+1}(x)$, $A_{r+2}(x), \cdots, A_{\lambda-1}(x)$ are equal, a circumstance which may occur for special choice of $A(x)$ and $B(x)$. If the degrees of $A_i(x)$ and $Q_i(x)$ are α_i and γ_i respectively, then it is clear that

$$\alpha_i - \alpha_{i+1} > \gamma_{i+1} \geqq 0 \qquad (i = 0, 1, \cdots, r-1).$$

The modification when the initial value of N is less than λ is obvious, and will be left to the reader. The results stated in the beginning of the paper are thus established.

5. We shall conclude by showing how the polynomials $A_{\lambda-1}(x)$, $A_{\lambda-2}(x)$, $\cdots, A_0(x)$ may be determined. We first observe that since

$$A(x)A_i(x) \equiv 0 \qquad (\text{modd } p^{\lambda-i}, B(x)) \tag{5.1}$$

we have $A(x)A_i(x) \equiv 0$ (modd $p^{\lambda-i-1}$, $B(x)$). Therefore by the fundamental property of $A_{i+1}(x)$,

$$A_i(x) \equiv 0 \qquad (\text{modd } p, A_{i+1}(x)) \qquad (i = 0, \cdots, \lambda - 1). \tag{5.2}$$

We have seen in §2 that we may assume that $B(x)$ is of the form $\{\phi(x)\}^\beta + pV(x)$ where $\phi(x)$ is primary and irreducible modulo p. If we construct a Schönemann decomposition of $A(x)$ modulo p^N, it is easily seen that we may assume that $A(x)$ is of the same form; thus

$$B(x) = \{\phi(x)\}^\beta + pV(x), \quad A(x) = \{\phi(x)\}^\alpha + pR(x) \tag{5.3}$$

where $\alpha < \beta$, and the degrees of $V(x)$ and $R(x)$ are less than those of $B(x)$ and $A(x)$ respectively. Hence

$$A(x)\{\phi(x)\}^{\beta-\alpha} \equiv pR(x)\{\phi(x)\}^{\beta-\alpha} - pV(x) \qquad (\text{mod } B(x)).$$

If p^M is the highest power of p dividing the right side of this last congruence, we have

$$A(x)\{\phi(x)\}^{\beta-\alpha} \equiv 0 \quad (\text{modd } p^M, B(x)), \quad \not\equiv 0 \quad (\text{modd } p^{M+1}, B(x))$$

and we may take

$$A_{\lambda-1}(x) = A_{\lambda-2}(x) = \cdots = A_{\lambda-M}(x),\ A_{\lambda-M}(x) \equiv \{\phi(x)\}^{\beta-\alpha} \pmod{p^M}.$$

Let i denote an integer $\leqq \lambda - M$. Then

$$A(x)A_i(x) \equiv p^{\lambda-i}S_i(x) \pmod{B(x)}, \tag{5.4}$$

where $S_i(x)$ is of lesser degree than $B(x)$.

We may assume that $S_i(x)$ *is not divisible by p and is of lesser degree than* $A(x)$. For since $A_i(x)$ is determined only modulo $p^{\lambda-i}$, if $S_i(x) = pS_i'(x)$ we have

$$A(x)(A_i(x) + p^{\lambda-i}) \equiv p^{\lambda-i}(A(x) + pS_i'(x)) \pmod{B(x)}$$

and by (5.3), the polynomial multiplying $p^{\lambda-i}$ on the right is not divisible by p. In the same way, if $S_i(x) = Q(x)A(x) + S_i''(x)$ where $S_i''(x)$ is of lesser degree than $A(x)$, then $Q(x)$ is necessarily of lesser degree than $A_i(x)$ so that $A_i(x) + p^{\lambda-i}Q(x)$ is a primary polynomial such that

$$A(x)(A_i(x) + p^{\lambda-i}Q(x)) \equiv p^{\lambda-i}S_i''(x) \pmod{B(x)}.$$

If $A_i(x)$ is known, we can determine $A_{i-1}(x)$. For, by (5.2),

$$A_{i-1}(x) = Q(x)A_i(x) + pR(x)$$

where $Q(x)$ must be primary, and $R(x)$ of lesser degree than $A_i(x)$. By (5.4),

$$A(x)A_{i-1}(x) \equiv p^{\lambda-i}Q(x)S_i(x) + pR(x)A(x) \pmod{B(x)}.$$

Take $R(x) = p^{\lambda-i-1}$. Then

$$A(x)A_{i-1}(x) \equiv 0 \quad (\text{modd } p^{\lambda-i+1}, B(x))$$

when and only when

$$Q(x)S_i(x) + A(x) \equiv 0 \quad (\text{modd } p, B(x)),$$

that is, when and only when

$$Q(x)S_i(x) + \{\phi(x)\}^\alpha \equiv 0 \quad (\text{modd } p, \{\phi(x)\}^\beta).$$

Since $S_i(x)$ is known and is of lesser degree than $\{\phi(x)\}^\alpha$ and not divisible by p, there exists a primary polynomial $Q(x)$ uniquely determined modulo p which satisfies this congruence. $A_{i-1}(x)$ is now uniquely determined modulo $p^{\lambda-i+1}$ and may be modified so as to satisfy the conditions corresponding to those imposed upon $A_i(x)$ in (5.4).

The remaining polynomials $A_{\lambda-M-1}(x), \cdots, A_1(x), A_0(x)$ can therefore be calculated step by step, and our solution is completed.

California Institute of Technology,
Pasadena, Calif.

A CHARACTERIZATION OF THE CLOSED 2-CELL*

BY

HASSLER WHITNEY†

1. Introduction. A number of characterizations have been given of the simple closed surface.‡ The proofs involve considerable point set difficulties. We give here a characterization of the closed 2-cell, that is, a point set homeomorphic with a circle and its interior. The fundamental theorem is partly of a combinatorial and partly of a continuity nature. It reads

THEOREM I. *Let R be a continuous curve § containing the simple closed curve J, such that*

(1) *J is irreducibly homologous to zero in R, and*

(2) *If γ is an arc with just its two end points a and b on J, then $R-\gamma$ is not connected.*

Let R' and J' be defined similarly. Then R and R' are homeomorphic, with J corresponding with J'.

That R is a closed 2-cell then follows immediately from the following theorem. We note that J corresponds with the circle, that is, J is the boundary of R.

THEOREM II. *If I is a circle in the plane and S is I with its interior, then S and I satisfy the conditions prescribed for R and J in the above theorem.*

The exact meaning of Condition (1) of Theorem I is given in §4; a stronger condition is the following: For every $\epsilon>0$ and any two points a and b on J, there is a set of points a_{ij} in R, $1\leqq i\leqq m$, $1\leqq j\leqq n$, such that all points a_{1j} coincide with a, all points a_{mj} coincide with b, all points a_{i1} lie on one arc ab of J, all points a_{in} lie on the other arc ab of J, and‖

$$\rho(a_{ij}, a_{i+1,j}) < \epsilon, \quad \rho(a_{ij}, a_{i,j+1}) < \epsilon;$$

moreover, this does not hold in any proper subset of R containing J.

* Presented to the Society, October 31, 1931; received by the editors April 13, 1932.

† National Research Fellow.

‡ That is, a point set homeomorphic with the surface of a sphere. See L. Zippin, American Journal of Mathematics, vol. 52 (1931), pp. 331–350; these Transactions, vol. 31 (1929), pp. 744–770; C. Kuratowski, Fundamenta Mathematicae, vol. 13 (1929), pp. 307–318; also references in these papers.

§ See Lemma A.

‖ $\rho(p, q)$ = distance from p to q, or in general, distance between two point sets; $\delta(S)$ = diameter of S; $V_\epsilon(S)$ = those points p for which $\rho(p, S)<\epsilon$; $W_\epsilon(S)$ = those points p for which $\rho(p, S)\leqq\epsilon$.

Notations and preliminary theorems are given in §§2, 3 and 4; an outline of the proof of Theorem I will be found in §5. The Jordan and related theorems follow of course from the above theorems.

2. **Point set background.** Elementary properties of point sets we shall need may be found in Hausdorff, *Mengenlehre*, chapter VI. A continuous curve is a metric space which can be expressed as the continuous image of a closed line segment. An arc is the topological image of a closed line segment; a simple closed curve, the topological image of a circle.

Two fundamental lemmas are the following:

LEMMA A.* *A compact, connected and locally connected metric space is a continuous curve, and conversely.*

LEMMA B.† *A continuous curve is arcwise connected.*

That is, any two points p and q in the set are end points of an arc pq in the set. Using the definition of a continuous curve, it is easily seen that two continuous curves which have common points form a continuous curve.

From these lemmas we deduce the following known theorems.

LEMMA C. *Any continuum C of diameter $<\epsilon$ in a continuous curve R is contained in a continuous curve C' in R of diameter $<\epsilon$.*

Say $\delta(C)=\epsilon-\epsilon'$. R being the continuous image of a closed line segment, we can divide this segment into segments so small that the diameter of the image of each is $<\epsilon'$. We let C' be the union of all of these images which have points in common with C.

LEMMA D. *A continuous curve R is locally arcwise connected.*

That is, given a point p and an $\epsilon>0$, there is a $\delta>0$ such that if $q\subset V_\delta(p)$, then there is an arc pq in R of diameter $<\epsilon$. As R is locally connected, we can take δ so that if $q\subset V_\delta(p)$, there is a continuum C in R of diameter $<\epsilon$ containing p and q. The continuum C is contained in a continuous curve C' of diameter $<\epsilon$, and C' is arcwise connected; hence there is an arc $pq\subset C'\subset R$, and $\delta(pq)<\epsilon$.

R is of course uniformly locally arcwise connected, by the Borel Theorem.

LEMMA E. *A connected open subset R' of a continuous curve R is arcwise connected.*

---

* See G. T. Whyburn, *Concerning continuous images of the interval*, American Journal of Mathematics, vol. 53 (1931), pp. 670–674.

† See references in R. L. Moore, *Report on continuous curves*, Bulletin of the American Mathematical Society, vol. 29 (1923), p. 293, footnote (†).

If there are two points p and q in R' which are joined by no arc in R', let A contain p and all points of R' joined to p by an arc in R', and put $B=R'-A$; then there is no arc joining a point of A to a point of B in R'. As R' is connected, there is a point p' in one of these sets, say B, which is a limit point of points of the other set, A. As R' is open in R, $\rho(p', R-R')=\epsilon>0$. We can take q' in A so close to p' that there is an arc $p'q'$ in R of diameter $<\epsilon$. But then $p'q'\subset R'$, a contradiction.

Suppose R is connected, and $p\subset R$ is such a point that $R-p$ is not connected. Then p is called a cut point of R.

LEMMA F. *Let R be a continuous curve without a cut point. Then for every $\epsilon>0$ there is a $\delta>0$ such that if $\rho(q, p)\geqq\epsilon$ and $\rho(q', p)\geqq\epsilon$, then there is an arc qq' with no points in $V_\delta(p)$.*

Suppose the contrary. Then there are three sequences of points $\{p_n\}$, $\{q_n\}$, $\{q_n'\}$, approaching points p, q, q', respectively, with $\rho(q_n, p_n)\geqq\epsilon$, $\rho(q_n', p_n)\geqq\epsilon$, and such that for each n, any arc q_nq_n' must contain points in $V_{\delta_n}(p_n)$, where $\lim_{n\to\infty}\delta_n=0$. By Lemma D it is seen that for any n greater than some N there are arcs q_nq, $q_n'q'$, with no points in $V_{\epsilon/2}(p)$. It follows that any arc qq' must pass through p, contradicting Lemma E (as p is not a cut point).

3. **Combinatorial background.*** A *k-simplex*, or *abstract k-simplex*, is a set of k elements (say points) $a_1a_2\cdots a_k$. The order in which we write the points is immaterial. For $k=0$, 1 and 2 we use also the terms *vertex*, *segment* and *triangle* respectively. A *k-chain* is a set of k-simplexes, and is written as the sum of these simplexes. The *sum* (mod 2) of several k-chains is the k-chain containing those simplexes which occur in an odd number of the k-chains.

The *boundary K* of a *k-simplex L*, $k>0$, is the sum of all $(k-1)$-simplexes formed by dropping out one of the vertices of the simplex. We write $L\to K$. A 0-simplex has no boundary. Thus

$$a\to 0,\ ab\to a+b,\ abc\to ab+ac+bc.$$

The *boundary* of a k-chain is the sum (mod 2) of the boundaries of the simplexes of the chain. Thus

$$ab+bc+cd\to a+d,\ abc+bcd\to ab+ac+bd+cd.$$

Evidently *the boundary of a sum of several k-chains is the sum of the boundaries of the chains.* If a k-chain has no boundary, it is called a k-cycle. (Any 0-chain is a 0-cycle.) *The boundary of a k-chain* ($k>0$) *is a* $(k-1)$*-cycle.* This is evi-

* Compare L. Vietoris, *Über den höheren Zusammenhang kompakter Räume*, Mathematische Annalen, vol. 97 (1927), pp. 454–472.

dent if the k-chain is a k-simplex. The general case then follows from the last theorem.

LEMMA G. *If $K \rightarrow a+b$ is a 1-chain, then there is a chain of segments aa_1, $a_1a_2, \cdots, a_nb$ in K.*

For otherwise we could divide the segments of K into two groups $K_1 \supset a$ and $K_2 \supset b$, no two simplexes from different groups having a common vertex. But then $K_1 \rightarrow a$, $K_2 \rightarrow b$, which cannot be, as the boundary of any 1-chain contains an even number of vertices.

A 1-circuit is a 1-cycle of the form $a_1a_2, a_2a_3, \cdots, a_{n-1}a_n, a_na_1$, the vertices being distinct except as shown.

LEMMA H. *Any 1-cycle K is a sum of 1-circuits.*

If a_1a_n is a segment of K, then $K+a_1a_n \rightarrow a_1+a_n$, as $K \rightarrow 0$ and $a_1a_n \rightarrow a_1+a_n$. We can thus find a set of distinct segments and vertices $a_1a_2, \cdots, a_{n-1}a_n$ in $K+a_1a_n$ not containing a_1a_n. This with a_1a_n is a 1-circuit K_1. As $K_1 \rightarrow 0$, $K+K_1$ is a 1-cycle containing no segments of K_1, and it contains a 1-circuit K_2. Continuing, we find $K=K_1+K_2+\cdots+K_m$.

4. A k-chain K is said to *lie in a point set* R if each vertex of K is in R. Any vertex now has both a name and a position. Two vertices are distinct if their names are distinct, irrespective of whether they coincide in position or not. ϵ being a positive number, a k-simplex $K \subset R$ is called an *(ϵ, k)-simplex* in R if $\delta(K)<\epsilon$, i.e. if any two vertices of K are within ϵ of each other. A k-chain is an *(ϵ, k)-chain* if each of its simplexes is an (ϵ, k)-simplex. A k-cycle K in S is said to be *ϵ-homologous to zero* ($K \epsilon\!\sim 0$) in R if there is an $(\epsilon, k+1)$-chain L in R of which K is the boundary. *If $K_1 \epsilon\!\sim 0$ and $K_2 \epsilon\!\sim 0$, then K_1+K_2 $\epsilon\!\sim 0$.* We write also $K_1 \epsilon\!\sim K_2$ for $K_1+K_2 \epsilon\!\sim 0$. *If $K_1 \epsilon\!\sim K_2$ and $K_2 \epsilon\!\sim K_3$, then $K_1 \epsilon\!\sim K_3$.*

Suppose the closed set R contains the simple closed curve J. If for every $\epsilon>0$ there is a $\delta>0$ such that any $(\delta, 1)$-cycle on J is $\epsilon\!\sim 0$ in R, then we say that *$J \sim 0$ in R*. If J is ~ 0 in R but is not ~ 0 in any proper closed subset of R containing J, then we say that J is *irreducibly* ~ 0 in R.

LEMMA I. *Given a simple closed curve J, let us divide it into the arcs* $\overline{a_1a_2}$, $\overline{a_2a_3}, \cdots, \overline{a_{n-1}a_n}, \overline{a_na_1}$, each of diameter $<\epsilon/2$. Let δ be smaller than the distance between any two of these arcs which have no common points. Then if $K'=a_1a_2+a_2a_3+\cdots+a_na_1$ and K is any $(\delta, 1)$-cycle on J, K is either $\epsilon\!\sim 0$ or $\epsilon\!\sim K'$ on J.*

By Lemma H, K is a sum of 1-circuits $K_1, \cdots, K_m$. If we show that each

* Here, $\overline{a_1a_2}$ denotes an arc, and a_1a_2, a segment.

K_i is $\epsilon\sim\alpha_i K'$, $\alpha_i=0$ or 1, it will follow that $K=\sum K_i\ \epsilon\sim\sum\alpha_i K'=0$ or K' (depending on whether $\sum\alpha_i$ is even or odd), and the lemma will be proved.

Consider any $K_i=b_1b_2+b_2b_3+\cdots+b_sb_1$, say. If a vertex b_j of K_i does not lie on any point a_k, say $b_j\subset\overline{a_ka_{k+1}}$; add to K_i the boundary of the ϵ-triangles $b_{j-1}b_ja_k'+b_jb_{j+1}a_k'$, where a_k' is a new vertex lying on a_k. The result is an $(\epsilon, 1)$-circuit $K_i^{(1)}\ \epsilon\sim K_i$, the vertex b_j having been replaced by the vertex a_k'. Repeat the process till we have an $(\epsilon, 1)$-circuit $K''=c_1c_2+c_2c_3+\cdots+c_sc_1\ \epsilon\sim K_i$.

Now any two consecutive vertices c_j, c_{j+1} lie on the same or consecutive vertices of K'. Suppose c_j is on a_k and c_{j+2} is on a_{k+p}, $p\neq 2$ or -2. Then add the boundary of $c_jc_{j+1}c_{j+2}$, replacing the segments $c_jc_{j+1}+c_{j+1}c_{j+2}$ by the single segment c_jc_{j+2}. Continue till we arrive at a (possibly void) $(\epsilon, 1)$-circuit $K^*=d_1d_2+\cdots+d_rd_1\epsilon\sim K_i$. If d_1 lies on a_k, then d_{j+1} lies on $a_{k\pm j}$, where we put $n+p=p$, etc.

If K^* contains no segments, $K_i\ \epsilon\sim 0$. Otherwise, following the vertices $d_1, d_2, \cdots, d_r, d_1$ of K^*, we have gone around J p times say. Add to K^* the boundaries of all the $2r$ ϵ-triangles of the following sort. If d_j lies on a_k, and d_{j+1} on $a_{k\pm1}$, two of the triangles are $d_jd_{j+1}a_k$ and $d_{j+1}\,a_ka_{k\pm1}$. The result is an $(\epsilon, 1)$-cycle $pK'=0$ or K'. Thus $K_i\epsilon\sim 0$ or K', and the proof is complete.

An immediate consequence of this lemma is

Lemma J. *Let the simple closed curve J lie in the closed set R. If for every $\epsilon>0$ there is a 1-cycle K' in J as above described which is $\epsilon\sim 0$ in R, then $J\sim 0$ in R.*

Lemma K. *If γ is an arc, then for every $\epsilon>0$ there is a $\delta>0$ such that any $(\delta, 1)$-cycle on γ is $\epsilon\sim 0$ on γ.*

The proof below holds in fact if γ is a closed k-cell, any k. It is sufficient to prove it for the case that γ is a closed line segment, in which case we can take $\delta=\epsilon/2$.†

Let K be a $(\delta, 1)$-cycle on γ, let a_0b_0 be a segment of K, and say $\delta(\gamma)=\alpha$. Choose a fixed point p in γ, and an integer $n>\alpha/\delta$. Let the vertices $a_1, a_2, \cdots, a_{n-1}$ divide the segment a_0p into n equal parts, and similarly for the vertices $b_1, b_2, \cdots, b_{n-1}$. Add to K the boundaries of all triangles of the form $a_ia_{i+1}b_i$, $a_{i+1}b_ib_{i+1}$, $a_{n-1}b_{n-1}p$, and of all similar triangles corresponding to the other segments of K. The result is 0. As all the triangles employed are ϵ-triangles, $K\epsilon\sim 0$ in γ.

5. Outline of the proof of Theorem I. The proof runs as follows.

† The essential point in the proof below is that γ is convex: any two points of γ are end points of a line segment in γ. The proof is then easily extended to the case of any set homeomorphic with γ.

(a) In §6 we show how an arc γ can be drawn in R crossing J,† avoiding two given closed sets. $R-\gamma$ is not connected.

(b) In §7 we prove some lemmas. These show (§8) that $R-\gamma$ contains exactly two components A' and B'. If $A=A'+\gamma$, then A and its boundary curve J_A (which is γ plus a part of J) satisfy condition (1) of the theorem; similarly for $B=B'+\gamma$ and J_B. Further, A and B are continuous curves.

(c) In §9 it is shown that any arc in A (or B) crossing J_A (J_B) divides A (B). Thus A and J_A (B and J_B) satisfy all the conditions of the theorem. Hence we can cut up each set just as we cut up R, and can continue indefinitely.

(d) The object of §10 is to prove that R may be cut into pieces of arbitrarily small diameter.

(e) The homeomorphism between R and R' is now easily established. We cut R up indefinitely, and cut R' in a corresponding fashion. Any point p of R lies in a descending sequence of pieces; the corresponding sequence in R' determines a point p', which we let correspond to p.

We turn now to the detailed proof.

6. **An arc crossing** J. We prove here

Lemma L.‡ *Let the simple closed curve J be ~ 0 in the continuous curve R. Let c and d be two points of J, dividing J into the two arcs η_1 and η_2. If C and D are two closed sets in R containing c and d respectively, and $C\cdot D=0$, then there is an arc γ in R joining η_1 to η_2 which has no points in C or in D.*

Say $\rho(C, D)=3\epsilon$, and put $C'=W_\epsilon(C)$, $D'=W_\epsilon(D)$; then $\rho(C', D')=\epsilon$. Take σ so small that any two points in R within σ of each other are joined by an arc of diameter $<\epsilon$ (Lemma D). Take δ so small that any (δ, 1)-cycle on J is $\sigma\sim 0$ in R. Construct the (δ, 1)-cycle $K=cc_1+c_1c_2+\cdots+c_md+dd_1+d_1d_2+\cdots+d_nc$, $c_i\subset\eta_1$, $d_i\subset\eta_2$. There is a (σ, 2)-chain

$$L = L_C + L_D \rightarrow K$$

in R, where we let L_C contain all those triangles of L with vertices in C', and let L_D be the rest of L.

Say

$$L_C \rightarrow K_C = K_C' + K^*,$$

where we let K_C' contain all those segments of K_C which are also in K. As $L_C\subset V_\sigma(C')$, $K^*\cdot D'=0$. Define K_D' by the relation

† That is, γ lies in R, and has only its end points on J.

‡ Compare P. Urysohn, *Über Räume mit verschwindender erster Brouwerscher Zahl*, Proceedings, Amsterdam Akademie van Wetenschappen, vol. 31 (1928), pp. 808–810.

$$L_D \to K_D = K_D' + K^*.$$

Adding these relations gives L on the left, and hence K on the right:

$$K = K_C' + K_D'.$$

As all the segments of K_C' are in K, K_D' must contain just those segments of K not in K_C'; in particular, it contains no segments of K^*. Hence all the segments of K^* are present in $K_D'+K^*$, the boundary of L_D (i.e. none have canceled out with segments of K_D'). Hence, as $L_D\cdot C'=0$,

$$K^*\cdot C' = K^*\cdot D' = 0.$$

As K_C is the boundary of L_C, it is a 1-cycle; hence

$$K_C + cc_1 = K_C' + K^* + cc_1 \to c + c_1.$$

By Lemma G, K_C+cc_1 contains a chain of segments joining c_1 to c. Following this chain, let p_s be the first vertex in η_2, and p_0, the last vertex before p_s in η_1, and say $p_0p_1, p_1p_2, \cdots, p_{s-1}p_s$ are the segments in between. We shall show that these segments are in K^*. If $s>1$ this is obvious, as then $p_1, \cdots p_{s-1}$ exist and are not on J. Suppose $s=1$ and p_0p_1 is not in K^*; then it is in $K_C'+cc_1$. It could only be the segment cc_1. But cc_1 lies in K and not in K_D, hence it is in K_C; it is not in K^*, hence it is in $K_C+K^*=K_C'$, and therefore not in $K_C'+cc_1$. This proves the statement.

Now let $\overline{p_ip_{i+1}}$ be an arc of diameter $<\epsilon$ in R, $i=0, \cdots, s-1$. These arcs form a continuous curve, from which we can pick out an arc γ (Lemma B) joining η_1 to η_2; we can take γ so only its end points are on J. As $p_ip_{i+1}\subset K^*$ and $\delta(\overline{p_ip_{i+1}})<\epsilon$, γ has no points in C or in D, and the lemma is proved.

7. We prove three lemmas.

LEMMA M. *If $J\subset C$, $J\sim 0$ in $C+D$, and $C\cdot D=$ an arc γ, then $J\sim 0$ in C.*

Given an $\epsilon>0$, choose first ϵ_1 so small that any $(3\epsilon_1, 1)$-cycle on γ is $\epsilon\sim 0$ in γ (Lemma K). Take next $\epsilon_2<\epsilon_1$ so that if $p\subset D$ and $\rho(p, C)<\epsilon_2$, then $\rho(p, \gamma)<\epsilon_1$. (If $D_1=D-D\cdot V_{\epsilon_1}(\gamma)$, take $\epsilon_2<\rho(D_1, C)$.) Take finally $\delta<\epsilon_2$ so that any $(\delta, 1)$-cycle K on J is $\epsilon_2\sim 0$ in $C+D$; we shall show that $K\epsilon\sim 0$ in C.

Let $L\to K$ be an $(\epsilon_2, 2)$-chain in $C+D$. Take any vertex p of L in $D\cdot V_{\epsilon_2}(C)-\gamma$, and replace it by a vertex $p'\subset\gamma$, where $\rho(p, p')<\epsilon_1$. L is thus replaced by a $(3\epsilon_1, 2)$-chain L', in which each triangle lies wholly in either C or D. Moreover, $L'\to K$, as no vertices of K have been moved.

Put $L'=L_C+L_D$, where L_C contains those triangles of L' in C. Say

$$L_C \to K + K^*;\text{ then } L_D \to K^*.$$

K^* is a $(3\epsilon_1, 1)$-cycle lying in $C\cdot D=\gamma$; it bounds an $(\epsilon, 2)$-chain L^* in γ. Hence

$$L_C + L^* \rightarrow (K + K^*) + K^* = K.$$

L_C+L^* is an $(\epsilon, 2)$-chain in C, and the lemma is proved.

Lemma N. *Let $A \cdot B = \gamma$, an arc whose end points are a and b. Let the arcs α and β join a and b in A and B respectively, neither having any points other than a and b in common with γ. If $\alpha+\beta\sim 0$ in $A+B$, then $\alpha+\gamma\sim 0$ in A.*

Given an $\epsilon>0$, choose ϵ_1, ϵ_2 and δ as in the last lemma. Take $(\delta, 1)$-chains K_α, K_β and K_γ in α, β and γ respectively, each bounded by $a+b$; by Lemma J, it is sufficient to show that $K_\alpha+K_\gamma\ \epsilon\!\sim 0$ in A.

$K_\alpha+K_\beta$ bounds an $(\epsilon_2, 2)$-chain L in $A+B$; we move each vertex of L in $B\cdot V_{\epsilon_2}(A)-\gamma$ onto γ, giving a $(3\epsilon_1, 2)$-chain $L'\rightarrow K_\alpha+K_\beta'$. Say $L'=L_A+L_B$, where $L_A\subset A$, $L_B\subset B$. If $L_A\rightarrow K_\alpha+K^*$, then $L_B\rightarrow K_\beta'+K^*$, and $K^*\subset\gamma$. K^*+K_γ is a $(3\epsilon_1, 1)$-cycle on γ bounding an $(\epsilon, 2)$-chain L^* in γ. Hence $L_A+L^*\rightarrow K_\alpha+K_\gamma$ in A, completing the proof.

Lemma O. *Let α, β and γ be three arcs such that $\alpha\cdot\beta=\alpha\cdot\gamma=\beta\cdot\gamma=a+b$. Say $\alpha+\gamma\subset A$ and $\beta+\gamma\subset B$. If $\alpha+\gamma\sim 0$ in A and $\beta+\gamma\sim 0$ in B, then $\alpha+\beta\sim 0$ in $A+B$.*

Define K_α, K_β, K_γ as before; we need merely show that $K_\alpha+K_\beta\,\epsilon\!\sim 0$ in $A+B$. There are $(\epsilon, 2)$-chains L_A and L_B such that $L_A\rightarrow K_\alpha+K_\gamma$ in A and $L_B\rightarrow K_\beta+K_\gamma$ in B; hence $L_A+L_B\rightarrow K_\alpha+K_\beta$ in $A+B$.

8. **The set $R-\gamma$.** Let γ be any arc in R crossing J; say the end points of γ divide J into the two arcs α and β. By condition (2) of the theorem, $R-\gamma$ is not connected. Let A' and B' be those components of $R-\gamma$ containing $<\alpha>$† and $<\beta>$ respectively. These are not the same component. For if they were, putting $A=A'+\gamma, D=R-A'$, we have $J\subset A$, $J\sim 0$ in $R=A+D$, and $A\cdot D=\gamma$; hence, by Lemma M, $J\sim 0$ in A, a proper subset of R, contrary to condition (1) of the theorem.

The same reasoning shows that R has no cut point p; we need merely replace γ by p in Lemma M and above.

Put

$$A = A' + \gamma,\ B = B' + \gamma.$$

If $D=R-A'$, then $A\cdot D=\gamma$ and $J=\alpha+\beta\sim 0$ in $R=A+D$. Hence, by Lemma N, $\alpha+\gamma\sim 0$ in A. Similarly, $\beta+\gamma\sim 0$ in B. Consequently, by Lemma O, $J\sim 0$ in $A+B$, from which follows that $A+B=R$.

Moreover, $\alpha+\gamma$ is irreducibly ~ 0 in A. For if $\alpha+\gamma\sim 0$ in A^*, $\alpha+\gamma\subset A^*\subset A$, then, by Lemma O, $\alpha+\beta\sim 0$ in A^*+B; hence $A^*+B=R$, which is only possible if $A^*=A$. Similarly, $\beta+\gamma$ is irreducibly ~ 0 in B.

† $<\alpha>$ is α except for its end points, etc.

Let us show that A is a continuous curve. It is connected, as A' is; it is self-compact, being a closed subset of a compact space. A is locally connected. For if p and q are points of A close enough together, there is an arc pq in R of small diameter; if pq lies partly in B', we can replace that part of it by an arc of γ of small diameter. Lemma A now applies. Similarly, B is a continuous curve.

9. We shall now show that any arc δ crossing $J_A=\alpha+\gamma$ in A divides A. The following two lemmas will be useful.

LEMMA P. *If η_1 and η_2 are arcs contained within the arcs γ and β respectively, then there is an arc pq crossing $J_B=\beta+\gamma$ in B, with $p\subset\eta_1$, $q\subset\eta_2$.*

This is an immediate consequence of Lemma L, if we take, for the closed sets of that lemma, the closed intervals of J_B complementary to η_1 and η_2.

LEMMA Q. *There are no two arcs ab and cd in R without common points, each crossing J, whose end points are in the order $acbd$ on J.*

This follows directly from what we have seen above.

To show that δ divides A, we must consider four cases.

Case 1. Both end points of δ lie on α. Suppose $A-\delta$ is connected; then it is arcwise connected, by Lemma E. Hence there is an arc in $A-\delta$ joining a point p of α lying between the two end points of δ and a point q within γ. If η_1 is an arc within γ containing q, there is an arc rs in B joining η_1 to a point s within β, with only its end points r and s on J_B, by Lemma P. The arc $pqrs$ crosses J and does not touch δ. But the end points of this arc alternate with those of δ on J, contradicting Lemma Q.

Case 2. δ is an arc cd, where c lies within α, d lies within γ. If $A-\delta$ is connected, let pq be an arc in this set joining points of α on opposite sides of c. If η_1 is an arc of γ containing d but not touching pq, let the arc rs join η_1 to β in B; then the arcs pq and $cdrs$ contradict Lemma Q.

Case 3. The end points c and d of δ lie within $\gamma=ab$, say in the order $acdb$. If $A-\delta$ is connected, let pq be an arc in this set joining a point p within α to a point q in γ between c and d. If η_1 is an arc of γ containing q but not touching δ, let r_1s_1 be an arc in B joining η_1 to a point s_1 within β.

The arcs acr_1 of γ and r_1s_1 form an arc acr_1s_1 crossing J; hence

$$R - acr_1s_1 = C_1 + C_2,$$

where C_1 contains the open arc $<as_1>$ of β, and C_2 contains b and points connected with b. As r_1s_1 lies in B, $A'\subset C_1+C_2$; the connected set $A'+b$ lies thus in C_2. If η_2 is an arc of γ containing c but not touching η_1, and r_2s_2 is an arc in C_1 joining η_2 to a point s_2 of β between a and s_1, then $\eta_2+r_2s_2$ does not touch pqr_1s_1, and has only the point c in common with δ.

Similarly, if η_3 is an arc of γ containing d but not touching η_1, there is an arc r_3s_3 in $R-bdr_1s_1$ such that r_3 lies in η_3, s_3 lies in β between s_1 and b, and $\eta_3+r_3s_3$ does not touch pqr_1s_1 and has only the point d in common with δ. The arc r_3s_3 does not touch r_2s_2, as it lies in C_2. Thus the two arcs pqr_1s_1 and $s_2r_2cdr_3s_3$ $(cd=\delta)$ contradict Lemma Q.

Case 4. The same as Case 3, except that $c=a$ or $d=b$, say the latter. Then, in the notation of Case 3, the arcs pqr_1s_1 and s_2r_2cb $(cb=\delta)$ contradict Lemma Q.

This completes the proof that A and J_A (B and J_B) satisfy the conditions of Theorem I.

10. **The cutting up of** R. We are concerned with the following lemma.

LEMMA R. *R may be cut into a finite number of pieces of arbitrarily small diameter.*

Given an $\epsilon>0$, choose $\delta<\epsilon$ so as to satisfy the requirement in Lemma F. Suppose R is cut up so that the diameter of the boundary of each piece is $<\delta$. Then each piece is of diameter $<3\epsilon$. For otherwise there is a point q of some piece R_i at a distance $\geqq\epsilon$ from its boundary J_i. Let p be a point of J_i, and q', a point of $R-R_i$ at a distance $\geqq\epsilon$ from p. Every arc from q to q' must cut the boundary J_i of R_i and thus must pass within δ of p, contradicting Lemma F.

The lemma thus follows from

LEMMA S. *Given a $\delta>0$, R can be cut up so that the diameter of the boundary of each piece is $<\delta$.*

Express R as the union of a finite number of continua:

$$R = K_1 + K_2 + \cdots + K_m,\ \delta(K_i) < \delta/2.$$

We shall cut up R in such a manner that no two of these continua K_i and K_j have points on the boundary of the same piece of R, if $K_i\cdot K_j=0$; the lemma will then follow.

Suppose we have cut R up a certain amount (perhaps not yet at all), into the pieces $R_1, R_2, \cdots, R_n$, with boundaries $J_1, J_2, \cdots, J_n$ (we may have R and J alone). Of course *each boundary J_i separates R_i from the rest of R*. Take any two continua, say K_1 and K_2, with $K_1\cdot K_2=0$, each of which has points on one of these J_i, say J_1. We shall cut R up further so that in the new pieces there is no one (i.e. no *piece*, not merely no *boundary* of a piece) which has any points in common with both K_1 and K_2; then on any further cutting up of R, this will still be true.

Divide the points of J_1 into three sets, as follows. We put a point x into the first set if it lies in K_1, or if following J_1 in both directions we reach points

of K_1 before reaching points of K_2; we put x into the second set if the same conditions hold with K_1 and K_2 interchanged; all other points we put into the third set. This set L_3' consists of open intervals of J_1, each being bounded by a point of K_1 on one end and a point of K_2 on the other. The points of the first set together with the points $K_1 \cdot R_1$ form a closed set L_1, and those of the second set together with $K_2 \cdot R_1$ form a closed set L_2. Then $\rho(L_1, L_2) > 0$ as $L_1 \cdot L_2 = 0$, from which follows that there are but a finite number of intervals in L_3'. As K_1 is connected, each component of L_1 has points on J_1, and thus on one of the intervals L_3 of J_1 complementary to the intervals of L_3'. Thus there are a finite number of components $L_{11}, L_{12}, \cdots, L_{1m_1}$ in L_1. Similarly there are a finite number of components $L_{21}, L_{22}, \cdots, L_{2m_2}$ in L_2.

We shall now cut R_1 into a number of pieces, in each of which either K_1 has no points or K_2 has no points. Suppose $L_{31}', \cdots, L_{3m_3}'$ and $L_{31}, \cdots, L_{3m_3}$ are the intervals of L_3' and L_3 respectively, and say they lie in the order L_{31}, L_{31}', L_{32}, L_{32}', $\cdots$, L_{3m_3}, L_{3m_3}' on J_1. If we go around J_1, the intervals of L_3 lie alternately in L_1 and L_2. Starting at L_{31}, which lies in L_{11} say, go around J_1 till we reach another interval L_{3k} in L_{11} (we may have gotten back to L_{31}). Put L_{32}, $L_{3,k-1}$ and all of J_1 between these into a set M_2' (which may be L_{32} alone), and put L_{3k}, L_{31}, and all of J_1 between these on the other side from L_{32} into a set M_1' (which may be L_{31} alone). L_{31}' and $L_{3,k-1}'$ are the two intervals of J_1 complementary to M_1' and M_2'.

No set L_{1i} or L_{2j} has points in both M_1' and M_2'. This follows for L_{11} by construction. If it were false for some other set, say L_{1s}, then L_{1s} would have points on two intervals L_{3p} and L_{3q} separated by L_{31} and L_{3k} on J_1. Now $L_{11} \cdot L_{1s} = 0$, hence $\rho(L_{11}, L_{1s}) > 0$. As R_1 is a continuous curve, there are continuous curves L_{11}^* and L_{1s}^* in R_1 containing L_{11} and L_{1s} and such that $L_{11}^* \cdot L_{1s}^* = 0$ (see Lemma C). These sets are arcwise connected, and we can draw arcs contradicting Lemma Q.

Let M_1 be M_1' plus all components L_{1i} and L_{2j} containing points of M_1', and define M_2 similarly. Then M_1 and M_2 are closed, $M_1 \cdot M_2 = 0$, and $M_1 + M_2 \supset L_1 + L_2$. By Lemma L we can draw an arc γ_1 from L_{31}' to $L_{3,k-1}'$ which has no points in M_1 or in M_2. R_1 is thus cut into two pieces, in each of which there is at least one component L_{1i} or L_{2j}; for one contains L_{11}, and the other contains that L_{2j} containing L_{32}. Thus in each piece there are less than $m_1 + m_2$ components, the number in R_1.

If one of the resulting pieces contains more than one component, we cut it up, etc. Finally each new piece of R_1 has points of only one component, and thus K_1 and K_2 are separated in R_1. We now separate K_1 and K_2 in each other piece R_i of R also. This is possible, for if K_i $(i = 1, 2)$ has points in any R_k, it also has points on J_k.

If now there are any other two of the continua K_i and K_j, $K_i\cdot K_j=0$, each of which has points on some new J_k, we cut R further till this is no longer true, etc. This completes the proof.

11. **The homeomorphism.** Cut R into pieces of diameter $<$ some σ. We make corresponding cuts in R' as follows. The first arc γ drawn in R cuts R into the two pieces R_1 and R_2 with boundaries J_1 and J_2 say. Draw any arc γ' crossing J' in R', cutting R' into the pieces R_1' and R_2' with boundaries J_1' and J_2'. We note that $J_1'+J_2'$ is homeomorphic with J_1+J_2, with J_k' corresponding to J_k, $k=1, 2$. Say γ_1 is an arc in R_1, cutting R_1 into pieces R_{11} and R_{12} with boundaries J_{11} and J_{12}. If a_1 and b_1 are the end points of γ_1, let a_1^* and a_2^* be the corresponding points of J_1' in the above homeomorphism. Draw an arc γ_1' crossing J_1' in R_1', with end points a_1' and b_1' close to a_1^* and b_1^* respectively (Lemma P); R_1' is divided thereby into the pieces R_{11}' and R_{12}' with boundaries J_{11}' and J_{12}'. Moreover, $J_{11}'+J_{12}'+J_2'$ is homeomorphic with $J_{11}+J_{12}+J_2$, with boundaries with the same subscripts corresponding.

In general, suppose $R_{i_1i_2\cdots i_m}$ is a piece that is present after R is cut a certain amount, and say the arc $\gamma_{i_1\cdots i_m}$ divides this set into the pieces $R_{i_1\cdots i_m1}$ and $R_{i_1\cdots i_m2}$, with boundaries $J_{i_1\cdots i_m1}$ and $J_{i_1\cdots i_m2}$. If $a_{i_1\cdots i_m}$ and $b_{i_1\cdots i_m}$ are the end points of $\gamma_{i_1\cdots i_m}$, let $a_{i_1}{}^*{}_{\cdots i_m}$ and $b_{i_1}{}^*{}_{\cdots i_m}$ be the corresponding points on $J_{i_1\cdots i_m}'$ in the homeomorphism we have already. Draw an arc $\gamma_{i_1}'{}_{\cdots i_m}$ crossing $J_{i_1}'{}_{\cdots i_m}$, with end points $a_{i_1}'{}_{\cdots i_m}$ and $b_{i_1}'{}_{\cdots i_m}$ close to the above points, dividing $R_{i_1}'{}_{\cdots i_m}$ into the pieces $R_{i_1}'{}_{\cdots i_m1}$ and $R_{i_1}'{}_{\cdots i_m2}$, with boundaries $J_{i_1}'{}_{\cdots i_m1}$ and $J_{i_1}'{}_{\cdots i_m2}$. The set of boundaries with primes is now homeomorphic with the set of boundaries without primes, boundaries with the same subscripts corresponding. We note that *if $R_{i_1\cdots i_m}$ and $R_{j_1\cdots j_m}$ have common points, then $R_{i_1}'{}_{\cdots i_m}$ and $R_{j_1}'{}_{\cdots j_m}$ have common points, and conversely.*

Having cut R into pieces of diameter $<\sigma$ and having cut R' in a corresponding fashion, we now cut each piece of R' into pieces of diameter $<\sigma/2$ and cut each piece of R in a corresponding fashion. Next we cut each resulting piece of R into pieces of diameter $<\sigma/4$, etc. Now for any $\epsilon>0$ there is an m such that

$$\delta(R_{i_1\cdots i_m}) < \epsilon, \quad \delta(R'_{i_1\cdots i_m}) < \epsilon,$$

for any m-fold subscript.

We now establish the homeomorphism between R and R'. Let p be any point of R. It lies in either R_1 or R_2 (perhaps in both), say in R_{i_1}. Then it lies in either R_{i_11} or R_{i_12} (perhaps in both), say in $R_{i_1i_2}$, etc. Thus we have a sequence of pieces

$$R \supset R_{i_1} \supset R_{i_1 i_2} \supset \cdots \supset p.$$

The corresponding pieces in R' have a single limit point:

$$R' \supset R'_{i_1} \supset R'_{i_1 i_2} \supset \cdots \supset p'.$$

This point p' we let correspond to p.

If there are different sequences of pieces in R containing p, we have different sequences in R' defining points p'. However, all these points p' are the same. For if R, R_{i_1}, $R_{i_1 i_2}$, $\cdots$, and R, R_{j_1}, $R_{j_1 j_2}$, $\cdots$, are two sequences containing p, then each piece $R_{i_1 \cdots i_m}$ has points in common with $R_{j_1 \cdots j_m}$, namely, the point p; hence, as we saw above, $R'_{i_1 \cdots i_m}$ and $R'_{j_1 \cdots j_m}$ have common points. Thus the corresponding sequences in R' close down on a single point. Similarly, to each point p' in R' corresponds a single point p in R.

Finally, the correspondence is continuous. For take a point p in R and an $\epsilon > 0$. Let p' be the corresponding point in R', and choose an m so that $\delta(R'_{i_1 \cdots i_m}) < \epsilon$ for all m-fold subscripts. Consider all the $R_{i_1 \cdots i_m}$ with m-fold subscripts which contain p; these include all points of R in some $V_\delta(p)$. Then if $q \subset V_\delta(p)$, the corresponding point q' is in $V_\epsilon(p')$, and the continuity is established. This completes the proof of Theorem I.

12. **Proof of Theorem II.** Let I be a circle in the plane, and let S be I plus its interior. S is self-compact, connected and locally connected, and is thus a continuous curve. That $I \sim 0$ in S follows from Lemma K.†

To show that I is irreducibly ~ 0 in S, suppose that $I \sim 0$ in S', a proper closed subset of S; we can suppose that S' is a continuous curve. Let p be a point of S not in S', and let $V_\delta(p)$ have no points in S'. Let ab be a segment of a straight line passing through p with its ends on I. Let a_1b_1 and a_2b_2 be parallel segments enclosing ab, and lying at a distance δ from ab. Then in that portion of S' between a_1b_1 and a_2b_2, the (short) arcs a_1a_2 and b_1b_2 are not connected. But if C and D are those parts of S' outside a_1b_1 and a_2b_2, by Lemma L we can draw an arc joining a_1a_2 to b_1b_2 in $S'-(C+D)$, a contradiction.

Finally, that an arc crossing I in S divides S is a special (and easily proved) case of the Jordan theorem. This completes the proof.

13. **The Jordan theorem.** Let J be a simple closed curve in the plane. Let I be a circle containing J in its interior. Draw two non-intersecting line segments from I to J. $S=I$ plus its interior is thus cut into three closed 2-cells, one of which, say R, has the boundary J. Then $R-J$ is the inside of J. The points of J are obviously accessible from either side.

† For S is a closed 2-cell.

PRINCETON UNIVERSITY,
PRINCETON, N. J.

NEW SETS OF INDEPENDENT POSTULATES FOR THE ALGEBRA OF LOGIC, WITH SPECIAL REFERENCE TO WHITEHEAD AND RUSSELL'S PRINCIPIA MATHEMATICA*

BY

EDWARD V. HUNTINGTON

INTRODUCTION

Three sets of independent postulates for the algebra of logic, or Boolean algebra, were published by the present writer in 1904. The *first set*, based on the treatment in Whitehead's *Universal Algebra*, is expressed in terms of $(K, +, \times)$, where K is a class of undefined elements, $a, b, c, \cdots$, and $a+b$ and $a\times b$ are the results of two undefined binary operations. The *second set* is expressed in terms of $(K, <)$, where $a<b$ is an undefined binary relation be-

---

* Presented to the Society, December 28, 1931, and September 2 and October 29, 1932; received by the editors June 27, 1932. A brief bibliography of postulates for Boolean algebra, which makes no pretence of being complete, is as follows:

E. Schröder, *Algebra der Logik*. Leipzig, Teubner, 1890.

A. N. Whitehead, *Universal Algebra*. Cambridge University Press, 1898.

E. V. Huntington, *Sets of independent postulates for the algebra of logic*. These Transactions, vol. 5 (1904), pp. 288–309.

E. Schröder, *Abriss der Algebra der Logik*. Leipzig, Teubner, 1909–1910.

A. Del Re, *Sulla indipendenza dei postulati della logica*. Rendiconto, Accademia delle Scienze, Naples, (3), vol. 17 (1911), pp. 450–458.

H. M. Sheffer, *A set of five independent postulates for Boolean algebra, with application to logical constants*. These Transactions, vol. 14 (1913), pp. 481–488.

B. A. Bernstein, *A complete set of postulates for the logic of classes expressed in terms of the operation "exception," and a proof of the independence of a set of postulates due to Del Re*. University of California Publications on Mathematics, vol. 1 (1914), pp. 87–96.

L. L. Dines, *Complete existential theory of Sheffer's postulates for Boolean algebras*. Bulletin of the American Mathematical Society, vol. 21 (1915), pp. 183–188.

B. A. Bernstein, *A set of four independent postulates for Boolean algebra*. These Transactions, vol. 17 (1916), pp. 50–51.

B. A. Bernstein, *A simplification of the Whitehead-Huntington set of postulates for Boolean algebras*. Bulletin of the American Mathematical Society, vol. 22 (1916), pp. 458–459.

J. G. P. Nicod, *A reduction in the number of the primitive propositions of logic*. Proceedings of the Cambridge Philosophical Society, vol. 19 (1917), pp. 32–41.

N. Wiener, *Certain formal invariances in Boolean algebras*. These Transactions, vol. 18 (1917), pp. 65–72.

C. I. Lewis, *A Survey of Symbolic Logic*. University of California Press, 1918.

H. M. Sheffer, *Review of C. I. Lewis's "A Survey of Symbolic Logic."* American Mathematical Monthly, vol. 27 (1920), pp. 309–311.

tween the elements a and b. The *third set* is expressed in terms of $(K, +)$, or, if one prefers, in terms of $(K, \times)$.

If the class K is finite, it is well known that the number of elements must be some power of 2; and any class consisting of 2, 4, 8, 16, $\cdots$ elements can be made into a Boolean algebra by properly defining $+$ and $\times$.

Every Boolean algebra contains a "zero element," z, such that $a+z=a$, and a "universe element," v, such that $a\times v=a$; and each element a determines an element a', called the "negative" of a, such that $a+a'=v$ and $a\times a'=z$.

In 1913, H. M. Sheffer published a set of postulates for the same algebra expressed in terms of $(K, |)$, where the "stroke," $|$, represents another binary operation, called "rejection," such that $a|b=(a+b)'$.

A. N. Whitehead and B. Russell, *Principia Mathematica*, second edition. Cambridge University Press, vol. 1, 1925.

H. M. Sheffer, *Review of "Principia Mathematica."* Isis, Quarterly organ of the History of Science Society, vol. 8(I) (1926), pp. 226–231.

Paul Bernays, *Axiomatische Untersuchung des Aussagen-Kalküls der "Principia Mathematica."* Mathematische Zeitschrift, vol. 25 (1926), pp. 305–320.

B. A. Bernstein, *Sets of postulates for the logic of propositions.* These Transactions, vol. 28 (1926), pp. 472–478.

D. Hilbert and W. Ackermann, *Grundzüge der theoretischen Logik.* Berlin, 1928.

Alfred Tarski, *Fundamentale Begriffe der Methodologie der deduktiven Wissenschaften.* I. Monatshefte für Mathematik und Physik, vol. 37 (1930), pp. 1–44.

Kurt Gödel, *Die Vollständigkeit der Axiome des logischen Funktionenkalküls.* Monatshefte, vol. 37 (1930), pp. 349–360.

Kurt Gödel, *Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme.* I. Monatshefte, vol. 38 (1931), pp. 173–198.

J. Lukasiewicz and A. Tarski, *Untersuchungen über den Aussagenkalkül.* Comptes Rendus des Séances de la Société des Sciences et des Lettres de Varsovie, vol. 23 (1930), Class III, pp. 1–21.

J. Lukasiewicz, *Philosophische Bemerkungen zu mehrwertigen Systemen des Aussagenkalküls.* Ibid., pp. 51–77.

A. Heyting, *Die formalen Regeln der intuitistischen Logik.* Sitzungsberichte der preussischen Akademie der Wissenschaften (Berlin), Jahrgang 1930, Physikalisch-Mathematische Klasse, pp. 42–56; 57–71; 158–169.

B. A. Bernstein, *Whitehead and Russell's theory of deduction as a mathematical science.* Bulletin of the American Mathematical Society, vol. 37 (1931), pp. 480-488.

Jørgen Jørgensen, *A Treatise of Formal Logic.* Oxford University Press, 3 vols., 1931.

E. V. Huntington, *A new set of independent postulates for the algebra of logic with special reference to Whitehead and Russell's Principia Mathematica.* (This brief abstract includes the "fourth set" in the present paper and one other set of a different character.) Proceedings of the National Academy of Sciences, vol. 18 (1932), pp. 179–180.

P. Henle, *The independence of the postulates of logic.* Bulletin of the American Mathematical Society, vol. 38 (1932), pp. 409–414.

B. A. Bernstein, *On proposition* *4.78 *of Principia Mathematica*, Bulletin of the American Mathematical Society, vol. 38 (1932), pp. 388–391.

C. I. Lewis and C. H. Langford, *Symbolic Logic.* New York, The Century Company, 1932.

In 1914, B. A. Bernstein gave a set in terms of $(K, -)$, where the "$-$" represents another binary operation called "exception," such that $a-b = a \times b'$; and also a set in terms of $(K, \div)$, where the "$\div$" indicates a binary operation called "adjunction," such that $a \div b = a + b'$.

In the meantime, the primitive propositions of Section A of the *Principia Mathematica* (1910) were expressed in terms of a class called the class of "elementary propositions," a binary operation called "disjunction," and a unary operation called "negation"; and Bernstein has recently shown (June, 1931) how these primitive propositions can be expressed in abstract mathematical form in terms of $(K, +, ')$. Since the relation between the theory of the *Principia* and the theory of Boolean algebra has been the subject of some discussion, it becomes a matter of interest to construct a set of independent postulates for Boolean algebra explicitly in terms of $(K, +, ')$, for comparison with the *Principia*.

The present paper contains several such sets, numbered in such a way as to avoid confusion with the first, second, and third sets of 1904.

The *fourth set*, containing six postulates, appears to be the simplest and most "natural" of all the sets of postulates for Boolean algebra. It contains no "existence" postulate.

The *fifth set*, suggested by Sheffer's set of 1913, is shorter by one postulate, but appears decidedly more "artificial" than the fourth set.

The *sixth set* is modeled after the *Principia*-Bernstein set, with the addition of an extra postulate which proves to be necessary to make the list sufficient for Boolean algebra. This set also appears artificial and complicated in comparison with the fourth set.

All three of these sets are expressed in terms of $(K, +, ')$; but since in all these sets (following the usual mathematical custom) tacit use is made of the equality sign, it is more accurate to say that all these sets are expressed in terms of $(K, +, ', =)$.

In the present paper, the rules governing the use of the equality sign are listed in explicit form as Postulates A, B, C, D. Such an explicit statement of the postulates governing the sign $=$ is essential to any satisfactory comparison between Boolean algebra and the *Principia*.

For, an outstanding feature of the *Principia* is that no postulates for $=$ are presupposed. The primitive propositions of the *Principia* do not contain the equality sign, and the development of the theory proceeds without the use of Postulates A, B, C, D. Instead, a symbol $\equiv$ is introduced by definition, and Postulates A, B, C, D (with $\equiv$ written in place of $=$) are supposed to be deduced as theorems.

It appears, however, that the desired properties of the sign $\equiv$, as de-

scribed in the informal part of the *Principia*, cannot be rigorously deduced from the formal list of primitive propositions and the formal definition of $\equiv$ in the *Principia*, without the use of some additional postulates.

In Appendix I of the present paper, the connection between a Boolean system $(K, +, ', =)$ and the *Principia* system $(K, +, ', \equiv)$ is explained; and in Appendix II a revised list of primitive propositions for the *Principia* is given.

The resulting expression of the *Principia*'s system in strictly postulational form is believed to be free from the objections which might be raised against any formulation (like Bernstein's of June, 1931) which pre-supposes the use of the equality sign.

The new set of postulates for the *Principia* are shown to be "consistent" and "independent" by the same methods that apply to any other set of mathematical postulates.

The first set (1904)

For convenience of reference, the postulates of the "first set" for Boolean algebra, which are expressed in terms of K, $+$, $\times$, are here reproduced, in abbreviated form, with the original numbering. (The original $\wedge$, $\vee$, and $\bar{a}$ are here replaced by z, v, and a'; and the circles around the $+$ and $\times$ are omitted.)

Ia. *If a and b are in the class K, then $a+b$ is in the class K.*

Ib. *If a and b are in the class K, then ab is in the class K.*

IIa. *There is an element z such that $a+z=a$ for every element a.*

IIb. *There is an element v such that $av=a$ for every element a.*

IIIa. $a+b=b+a$.

IIIb. $ab=ba$.

IVa. $a+bc=(a+b)(a+c)$.

IVb. $a(b+c)=ab+ac$.

V. *For each element a there is an element a' such that $a+a'=v$ and $aa'=z$.*

VI. *There are at least two distinct elements in the class K.*

From these postulates the following theorems are deduced in the paper cited.

VIIa. *The z in* IIa *is unique.* VIIb. *The v in* IIb *is unique.*

VIIIa. $a+a=a$. VIIIb. $aa=a$.

IXa. $a+v=v$. IXb. $az=z$.

Xa. $a+ab=a$. Xb. $a(a+b)=a$.

XI. *The element a' in* V *is uniquely determined by a.*

XIIa. $a+b=(a'b')'$. XIIb. $ab=(a'+b')'$.

XIIIa. $(a+b)+c=a+(b+c)$. XIIIb. $(ab)c=a(bc)$.

XIV. *The relation* $a<b$ *is defined by any one of the following equations:*

$$a+b=b; \quad ab=a; \quad a'+b=v; \quad ab'=z.$$

Concerning the relation $<$ we have the following theorems, 2.1–2.9, which correspond to the postulates 1–9 of the "second set" in the paper of 1904.

2.1. $a<a$.

2.2. *If* $a<b$ *and* $b<a$, *then* $a=b$.

2.3. *If* $a<b$ *and* $b<c$, *then* $a<c$.

2.4. $z<a$ (*where* z *is the element in* IIa *and* VIIa).

2.5. $a<v$ (*where* v *is the element in* IIb *and* VIIb).

2.6. $a<a+b$; *and if* $a<y$ *and* $b<y$, *then* $a+b<y$.

2.7. $ab<a$; *and if* $x<a$ *and* $x<b$, *then* $x<ab$.

2.8. *If* $x<a$ *and* $x<a'$, *then* $x=z$; *and if* $a<y$ *and* $a'<y$, *then* $y=v$.

2.9. *If* $a<b'$ *is false, then there is at least one element* x, *distinct from* z, *such that* $x<a$ *and* $x<b$.

Examples of Boolean algebras*

The most familiar example of a Boolean algebra is the following:

K = the class of regions in a square (including the null region, and the whole square);

$a+b$ = the smallest region which includes both a and b;

a' = the region complementary to a with respect to the square;

ab = the region common to a and b.

Here the relation $a<b$ means "a is included in b."

Another interesting example is the following, given by H. M. Sheffer in his review of C. I. Lewis's *A Survey of Symbolic Logic* (American Mathematical Monthly, vol. 27 (1920), p. 310):

K = a class of eight numbers, 1, 2, 3, 5, 6, 10, 15, 30;

$a+b$ = the least common multiple of a and b;

$a'=30/a$;

ab = the highest common factor of a and b.

Here the relation $a<b$ means "a is a factor of b."

Or, in general, let K = the class of 2^m numbers which are the factors of any Boolean integer, v ("Boolean integer" being the name given by Sheffer to any integer which contains no square factor); with $a+b$, a', and ab defined as illustrated above for the case $v=30$.

Another example for eight elements is the following:

* The name Boolean algebra (or Boolean "algebras") for the calculus originated by Boole, extended by Schröder, and perfected by Whitehead seems to have been first suggested by Sheffer, in 1913.

K = a class of eight numbers: 0; 2, 3, 4; 23, 24, 34; and 234 $(=v)$; $a+b$, a' and ab being defined as in the accompanying tables.

+	0	234	2	34	3	24	4	23	′
0	0	234	2	34	3	24	4	23	234
234	234	234	234	234	234	234	234	234	0
2	2	234	2	234	23	24	24	23	34
34	34	234	234	34	34	234	34	234	2
3	3	234	23	34	3	234	34	23	24
24	24	234	24	234	234	24	24	234	3
4	4	234	24	34	34	24	4	234	23
23	23	234	23	234	23	234	234	23	4

×	0	234	2	34	3	24	4	23
0	0	0	0	0	0	0	0	0
234	0	234	2	34	3	24	4	23
2	0	2	2	0	0	2	0	2
34	0	34	0	34	3	4	4	3
3	0	3	0	3	3	0	0	3
24	0	24	2	4	0	24	4	2
4	0	4	0	4	0	4	4	0
23	0	23	2	3	3	2	0	23

It will be observed that the digits in $a+b$ include the digits in a and also the digits in b (0 not counting as a digit); and the digits in ab are the digits common to a and b. Hence the commutative, associative and distributive laws are seen at once to be true. Also, the numbers 0 and 234 are seen to serve as the elements z and v. By the same process, we can readily construct an example for 2^m elements, where m is any integer.

The tables for four elements are conveniently written as follows:

+	0	1	2	3	′
0	0	1	2	3	1
1	1	1	1	1	0
2	2	1	2	1	3
3	3	1	1	3	2

×	0	1	2	3
0	0	0	0	0
1	0	1	2	3
2	0	2	2	0
3	0	3	0	3

(These tables are the same as the upper left hand quarters of the tables for eight elements, the digit 4 being dropped, and the universe element 234 being represented by 1.)

Postulates governing the use of the equality sign

The postulates of the fourth, fifth, and sixth sets are expressed in terms of the undefined concepts (K, +, ′), the first two postulates in each set being the following:

Postulate 1. *If a and b are in the class K, then $a+b$ is in the class K;*

POSTULATE 2. *If a is in the class K, then a' is in the class K;*
and in each of these sets (following the usual mathematical procedure), the use of the equality sign, $=$, is taken for granted.

If preferred, however, the equality sign itself may be regarded as an additional undefined concept, provided suitable postulates are laid down governing its use.

An obvious set of postulates for $=$ is as follows, where $a, b, c, \cdots$ are understood to be elements of the class K.

POSTULATE A. *If a is in the class K, then $a=a$.*

POSTULATE B. *If $a=b$, then $b=a$.*

POSTULATE C. *If $a=b$ and $b=c$, then $a=c$.*

POSTULATE D. *If $x=y$, then $f(x, a, b, c, \cdots)=f(y, a, b, c, \cdots)$, where $f(x, a, b, c, \cdots)$ is any element of the class K built up from the elements $x, a, b, c, \cdots$ by successive applications of the operators $+$ and $'$ (see Postulates* 1 *and* 2), *and $f(y, a, b, c, \cdots)$ is the element obtained from $f(x, a, b, c, \cdots)$ by writing y in place of x throughout.*

If these postulates A, B, C, D are added, the fourth, fifth, and sixth sets of postulates may be said to express Boolean algebra in terms of the *four* undefined concepts $(K, +, ', =)$.

THE FOURTH SET

The following set of independent postulates for Boolean algebra is expressed in terms of $(K, +, ')$. K is a class of elements $a, b, c, \cdots$; $a+b$ denotes the result of a binary operation called logical addition; and a' denotes the result of a unary operation called logical negation. (A trivial preliminary postulate 4.0, demanding that the class K shall contain at least two distinct elements, is assumed without further mention; and in Postulates 4.3–4.6 it is assumed that the indicated combinations are elements of K. Also, Postulates A, B, C, D are assumed without further mention.)

POSTULATE 4.1. *If a and b are in the class K, then $a+b$ is in the class K.*

POSTULATE 4.2. *If a is in the class K, then a' is in the class K.*

POSTULATE 4.3. $a+b=b+a$.

POSTULATE 4.4. $(a+b)+c=a+(b+c)$.

POSTULATE 4.5. $a+a=a$.

POSTULATE 4.6. $(a'+b')'+(a'+b)'=a$.

By aid of the usual definition of $a\times b$ (or ab), namely:

4.7. **Definition.** $ab=(a'+b')'$,

the last postulate can be thrown into the following more familiar form:

4.8. $ab+ab'=a$.

From 4.6, by 4.3, we have $(a'+b)'+(a'+b')'=a$, whence by 4.2, $(a'+b')'+(a'+b'')'=a$. But by 4.7, $(a'+b')'=ab$ and $(a'+b'')'=ab'$. Hence $ab+ab'=a$. Conversely, from 4.7 and 4.8 we have $(a'+b')'+(a'+b'')'=a$, whence by 4.10, below, $(a'+b')'+(a'+b)'=a$.

The *consistency* of these postulates is established by the existence of any system $(K, +, ')$ which satisfies them all, as, for example, any one of the examples of Boolean algebra mentioned above.

To establish the *equivalence* of this fourth set (which is expressed in terms of $K, +, '$) and the first set of 1904 (which is expressed in terms of $K, +, \times$), we must show (1) that all the postulates of the fourth set are deducible from the postulates of the first set, when a' is properly defined in terms of $+$ and $\times$; and (2) that all the postulates of the first set are deducible from the postulates of the fourth set, when $a\times b$ is properly defined in terms of $+$ and $'$.

The first part of the proof is immediately evident from the preceding section.

The second part of the proof is provided by the following theorems which are deduced from Postulates 4.1–4.6, with the aid of the definition of $a\times b$ contained in 4.7.

4.9. $a+a'=a'+a''$.

By 4.6, $[a]+[a']=[(a'+a''')'+(a'+a'')']+[(a''+a''')'+(a''+a'')']$ and $[a']+[a'']=[(a''+a'')'+(a''+a')']+[(a'''+a'')'+(a'''+a')']$. Hence by 4.3 and 4.4, $a+a'=a'+a''$.

Alternative proof, using 4.7 and 4.8 in place of 4.6: By 4.8, $a+a'=(aa'+aa'')+(a'a'+a'a'')$ and $a'+a''=(a'a+a'a')+(a''a+a''a')$. Hence by 4.3 and 4.4 (since by 4.3 and 4.7, $ab=ba$), we have $a+a'=a'+a''$.

4.10. $a''=a$.

By 4.6, $(a'''+a'')'+(a'''+a')'=a''$ and $(a'+a''')'+(a'+a'')'=a$. But by 4.9, $a'+a''=a''+a'''$. Hence by 4.3, $a''=a$.

Alternative proof, using 4.7 and 4.8 in place of 4.6: By 4.8, $a''a+a''a'=a''$ and $aa'+aa''=a$. Hence by 4.7 and 4.3, $(a'+a''')'+(a''+a''')'=a''$ and $(a'+a'')'+(a'+a''')'=a$. But by 4.9, $a'+a''=a''+a'''$. Hence by 4.3, $a''=a$.

4.11. $a+a'=b+b'$.

Let $x=a+a'$ and $y=b+b'$. Then by 4.3, 4.6, 4.5, 4.4, 4.9, $y=b'+b=b'+[(b'+b')'+(b'+b)']=b'+(b''+y')=(b'+b'')+y'=(b+b')+y'=y+y'$. But by 4.6, with 4.3 and 4.4,

$$
\begin{aligned}
y+y' &= [(y'+x'')'+(y'+x')']+[(y''+x'')'+(y''+x')'] \\
&= [(x''+y'')'+(x''+y')']+[(x'+y'')'+(x'+y')'].
\end{aligned}
$$

Hence, by 4.6, $y=x'+x$.

Again, by 4.3, 4.6, 4.5, 4.4, 4.9,

$$x = a' + a = a' + [(a' + a')' + (a' + a)'] = a' + (a'' + x')$$
$$= (a' + a'') + x' = (a + a') + x' = x + x'.$$

Hence, by 4.3, $x=y$.

4.12. Definition. $v=a+a'=$the "universe" element of the system.

This element v exists, by 4.1 and 4.2, and is unique, by 4.11. Moreover, by 4.3, $v=a'+a$.

4.13. Definition. $z=(a+a')'=$the "zero" element of the system.

This element z exists, by 4.1 and 4.2, and is unique, by 4.11. Obviously, $z=v'$, where v is the universe element of 4.12; and by 4.10, $z'=v$.

4.14. *If* $a'=b'$, *then* $a=b$. (By 4.2 and 4.10.)

4.15. $z+a=a$.

By 4.6, $(a'+a')'+(a'+a)'=a$. Hence by 4.5 and 4.3, $(a')'+(a+a')'=a$. Hence by 4.10 and 4.12, $a+v'=a$, whence by 4.13 and 4.3, $z+a=a$.

4.16. $av=a$.

By 4.7, 4.3, 4.13, 4.15, 4.10, $av=(a'+v')'=(v'+a')'=(z+a')'=(a')'=a$.

4.17. $aa'=z$.

By 4.7, 4.12, 4.13, $aa'=(a'+a'')'=v'=z$.

4.18. $ab=ba$. (By 4.7 and 4.3.)

4.19. $(ab)c=a(bc)$.

By 4.7 and 4.10, $(ab)c=(a'+b')'c=[(a'+b')''+c']'=[(a'+b')+c']'$ and $a(bc)=a(b'+c')'=[a'+(b'+c')'']'=[a'+(b'+c')]'$. But these two values are equal by 4.4.

4.20. $a+b=(a'b')'$.

By 4.7 and 4.10, $a'b'=(a''+b'')'=(a+b)'$. Hence by 4.10, $(a'b')'=(a+b)''=a+b$.

4.21. $aa=a$.

By 4.7, 4.5, 4.10, $aa=(a'+a')'=(a')'=a$.

4.22. $a+v=v$.

By 4.12, 4.4, 4.5, 4.12, $a+v=a+(a+a')=(a+a)+a'=a+a'=v$.

4.23. $az=z$.

By 4.7, 4.13, 4.22, $az=(a'+z')'=(a'+v)'=v'=z$.

4.24. $a+ab=a$.

By 4.8, $ab+ab'=a$. Hence by 4.3, 4.4, 4.5, $a+ab=ab+a=ab+(ab+ab')=(ab+ab)+ab'=ab+ab'=a$.

4.25. $a(a+b)=a$.

By 4.7, 4.20, 4.10, 4.24, 4.10, $a(a+b)=[a'+(a+b)']'=[a'+(a'b')'']'=[a'+a'b']'=[a']'=a$.

4.26. *If* $a'+b=v$ *and* $b'+a=v$, *then* $a=b$.

By 4.15 and 4.3, $a+v'=a$. By 4.6, $(a'+b')'+(a'+b)'=a$. Hence if $a'+b=v$, $(a'+b')'=a$. By 4.6, $(b'+a')'+(b'+a)'=b$. Hence if $b'+a=v$, $(b'+a')'=b$. Hence by 4.3, $a=b$.

4.27. *If* $a+b=v$ *and* $ab=z$, *then* $a'=b$.

From $a+b=v$, by 4.10, $a''+b=v$. From $ab=z$, by 4.7, $(a'+b')'=z$, whence by 4.10, 4.13, 4.3, $b'+a'=v$. Hence by 4.26, $a'=b$.

In the following theorems, parentheses are omitted, in view of the associative laws, 4.4 and 4.19, and references to these laws, and to the commutative laws, 4.3 and 4.18, will often be understood.

4.28. $abc+abc'+ab'c+ab'c'+a'bc+a'bc'+a'b'c+a'b'c'=v$.

By 4.8, the given sum $=ab+ab'+a'b+a'b'=a+a'$, and by 4.12, $a+a'=v$.

4.29. *If* A *and* B *are any two distinct terms of the sum in* 4.28, *then* $AB=z$.

For example, $(ab'c)(a'bc)=(aa')(b'cbc)=z(b'cbc)=z$ by 4.17 and 4.23.

4.30. $ab+ac=abc+abc'+ab'c$.

By 4.8, $ab=abc+abc'$ and $ac=abc+ab'c$. Hence by 4.5, $ab+ac=abc+abc'+ab'c$.

4.31. $[a(b+c)]'=ab'c'+a'bc+a'bc'+a'b'c+a'b'c'$.

By 4.7, 4.10, $[a(b+c)]'=a'+(b+c)'=a'+b'c'$. But by 4.8 $a'=a'b+a'b'=a'bc+a'bc'+a'b'c+a'b'c'$, and by 4.18, 4.8, $b'c'=ab'c'+a'b'c'$. Hence the theorem, by 4.5.

4.32. $(ab+ac)+[a(b+c)]'=v$. (From 4.30, 4.31, by 4.28.)

4.33. $(ab+ac)[a(b+c)]'=z$.

Let A, B, C, D, E, F, G, H be the eight terms in 4.28. By 4.30, $ab+ac=A+B+C$, and by 4.31, $[a(b+c)]'=D+E+F+G+H$. By 4.29, 4.5, $AD+AE=z+z=z$. Hence by 4.32, 4.15, $[A(D+E)]'=v$, whence by 4.10, 4.13, $A(D+E)=z$. Similarly, $A(D+E)+AF=z+z=z$, whence $A(D+E+F)=z$. And so on; so that $A(D+E+F+G+H)=z$. By similar reasoning, we find $(A+B+C)(D+E+F+G+H)=z$, which proves the theorem.

4.34. $a(b+c)=ab+ac$. (First form of the distributive law.)

From 4.32 and 4.33, by 4.27, $(ab+ac)'=[a(b+c)]'$. Hence by 4.14, $ab+ac=a(b+c)$.

4.35. $a+bc=(a+b)(a+c)$. (Second form of the distributive law.)

By 4.10, 4.7, $a+bc=(a')'+(b+c')'=[a'(b'+c')]'$, whence by 4.34, $a+bc=[a'b'+a'c']'$. But also $(a+b)(a+c)=[(a+b)'+(a+c)']'=[a'b'+a'c']'$. Hence $a+bc=(a+b)(a+c)$.

These propositions include all the postulates of the first set of 1904 (see 4.1, 4.7, 4.15, 4.16, 4.3, 4.18, 4.35, 4.34, 4.12, 4.17); so that any system (K, $+$, $'$) which satisfies Postulates 4.1-4.6 will have all the properties of a

Boolean algebra, if the logical product, $a\times b$, is defined in terms of $+$ and $'$ as in 4.7.

To prepare the way for the definition of the relation $a<b$, we prove the following theorems.

4.36. *If* $a+b=b$, *then* $ab=a$; and conversely, *if* $ab=a$, *then* $a+b=b$.

If $a+b=b$, then by 4.7, 4.20, 4.24, 4.10,

$$ab = (a' + b')' = [a' + (a + b)']' = [a' + a'b']' = [a']' = a.$$

If $ab=a$, then by 4.20, 4.7, 4.10, 4.24,

$$a + b = (a'b')' = [(ab)'b']' = [(ab)'' + b'']'' = ab + b = b.$$

4.37. *If* $a+b=b$, *then* $a'+b=v$; and conversely, *if* $a'+b=v$, *then* $a+b=b$.

If $a+b=b$, then by 4.7 and 4.22, $a'+b=a'+(a+b)=(a'+a)+b=v+b=v$. If $a'+b=v$, then by 4.20, 4.15, 4.17, 4.34, 4.7, 4.10, 4.15, $a+b=(a'b')'=[a'b'+z]'=[a'b'+bb']'=[(a'+b)b']'=(a'+b)'+b=v'+b=z+b=b$.

4.38. *If* $a+b=b$, *then* $ab'=z$; and conversely, *if* $ab'=z$, *then* $a+b=b$.

If $a+b=b$, then by 4.7, 4.12, 4.22,

$$ab' = (a' + b)' = [a' + (a + b)]' = [(a' + a) + b]' = (v + b)' = v' = z.$$

If $ab'=z$, then by 4.20, 4.10, $a'+b=(ab')'=z'=v$, whence by 4.37, $a+b=b$.

4.39. Definition. If $a+b=b$; or if $ab=a$; or if $a'+b=v$; or if $ab'=z$; then and only then we write $a<b$.

The equivalence of these four forms of the definition follows from 4.36, 4.37, 4.38.

The following theorems are added because of their connection with the fifth and sixth sets, below.

4.40. $a+(b+c)'=[(b'+a)'+(c'+a)']'$.

By 4.7, 4.10, 4.34, 4.21, 4.24, 4.7, $[(b'+a)'+(c'+a)']'=(b'+a)(c'+a)=b'c'+ab'+ac'+aa=[a+a(b'+c')]+b'c'=[a]+b'c'=a+(b+c)'$.

4.41. $(b'+c)'+[(a+b)'+a+c]=v$.

By 4.7, 4.20, 4.16, 4.12, 4.34, 4.35,

$$(b' + c)' = bc' = vbc' = (a + a')bc' = abc' + a'bc';$$

$$(a + b)' = a'b' = a'b'v = a'b'(c + c') = a'b'c + a'b'c';$$

$$\begin{aligned} a + c &= av + vc = a(b + b') + (b + b')c = ab + ab' + bc + b'c \\ &= (abc + abc') + (ab'c + ab'c') + (abc + a'bc) + (ab'c + a'b'c). \end{aligned}$$

Hence the theorem, by 4.28 (with 4.5).

4.42. $(a+a)'+a=v$. (By 4.5, 4.12.)

4.43. $b'+(a+b)=v$.

By 4.3, 4.4, 4.12, 4.22, $b'+(a+b)=a+(b+b')=a+v=v$.

4.44. $(a+b)'+(b+a)=v$. (By 4.3, 4.12.)

Independence proofs for the fourth set

The *independence* of the postulates of the fourth set is established by the existence of the following examples of systems $(K, +, ')$, each of which violates the like-numbered postulate, and satisfies all the other postulates of the set.

Example 4.1. K = two elements, 0 and 1, with $+$ and $'$ defined as follows: $0+0=0$, $1+1=1$, $0+1=x$, $1+0=x$; $0'=1$, $1'=0$. Here x is not an element of the class K, so that Postulate 4.1 fails. The other postulates are satisfied whenever the indicated combinations are elements of the class.

Example 4.2. K = two elements, 0 and 1, with $a+b$ and a' defined as in the accompanying table. Here x is not in the class K, so that Postulate 4.2 fails. The other postulates are satisfied whenever the indicated combinations are elements of the class.

+	0	1	′
0	0	1	x
1	1	1	x

Example 4.3. In this system, Postulate 4.3 fails, since $5+2=5$ and $2+5=2$. The other postulates will be found to be satisfied.

+	0	1	2	3	4	5	′
0	0	1	2	3	4	5	1
1	1	1	1	1	1	1	0
2	2	1	2	1	1	2	3
3	3	1	1	3	3	1	2
4	4	1	1	4	4	1	5
5	5	1	5	1	1	5	4

Example 4.4. Here 4.4 fails, since $(2+1)+3=2+3=1$ while $2+(1+3)=2+0=2$. The other postulates are satisfied.

+	0	1	2	3	′
0	0	1	2	3	1
1	1	1	2	0	0
2	2	2	2	1	3
3	3	0	1	3	2

Example 4.5. Here 4.5 fails since $2+2=1$. The other postulates are satisfied.

$+$	0	1	2	$'$
0	0	1	2	1
1	1	1	1	0
2	2	1	1	2

Example 4.6. Here 4.6 fails since $(3'+5')'+(3'+5)'=(2+4)'+(2+5)'=1'+1'=0+0=0\neq 3$. The other postulates are found to be satisfied.

$+$	0	1	2	3	4	5	$'$
0	0	1	2	3	4	5	1
1	1	1	1	1	1	1	0
2	2	1	2	1	1	1	3
3	3	1	1	3	1	1	2
4	4	1	1	1	4	1	5
5	5	1	1	1	1	5	4

The fifth set

The following set of independent postulates for Boolean algebra is directly suggested by H. M. Sheffer's postulates of 1913, when these are expressed in terms of $(K, +, ')$.

This fifth set contains one fewer postulate than the fourth set; but the fifth set as a whole seems less simple and natural than the fourth set.

(A trivial preliminary postulate 5.0 demanding that the class contain at least two elements is assumed without further mention, and in Postulates 5.3–5.5 it is assumed that the indicated combinations are elements of K. Also, Postulates A, B, C, D are assumed without further mention.)

Postulate 5.1. *If a and b are in the class K, then $a+b$ is in the class K.*

Postulate 5.2. *If a is in the class K, then a' is in the class K.*

Postulate 5.3. $(a')'=a$.

Postulate 5.4. $a+(b+b')'=a$.

Postulate 5.5. $a+(b+c)'=[(b'+a)'+(c'+a)']'$.

The *consistency* of these five postulates is shown by any example of a Boolean algebra $(K, +, ')$, like the regions within a square, with $+$ and $'$ defined in the usual way.

The *equivalence* between the fifth set and the earlier sets is established as follows.

5.6. $a+a=a$.

By 5.5 and 5.3, $b'+(b+b')'=[(b'+b')'+(b+b')']'$, whence, applying 5.4 to each side, $b'=[(b'+b')']'$. Hence by 5.3, $b'=b'+b'$. Hence $a''=a''+a''$, whence by 5.3, $a=a+a$.

5.7. $a+b=b+a$.

By 5.3, 5.6, 5.5, 5.6, 5.3,

$$a + b = a + (b')' = a + (b' + b')' = [(b'' + a)' + (b'' + a)']' = [(b'' + a)']' = b + a.$$

5.8. $(a'+b')'+(a'+b)'=a$.

By 5.4, 5.5, 5.3, 5.7,

$$a' = a' + (b + b')' = [(b' + a')' + (b'' + a')']' = [(b' + a')' + (b + a')']' = [(a' + b')' + (a' + b)']'.$$

Hence by 5.3, $a=(a'+b')'+(a'+b)'$.

The following theorems lead up to the associative law, 5.27.

5.9. $a+a'=b+b'$.

By 5.4, 5.7, 5.4,

$$(a + a')' = (a + a')' + (b + b')' = (b + b')' + (a + a')' = (b + b')'.$$

Hence by 5.3, $a+a'=b+b'$.

5.10. Definition. $v=a+a'=$the "universe" element of the system.

This element v exists, by 5.1 and 5.2, and is unique by 5.9. Moreover, by 5.7, $v=a'+a$.

5.11. Definition. $z=(a+a')'=$the "zero" element of the system.

This element z exists, by 5.1 and 5.2, and is unique by 5.9. Obviously, $z=v'$, where $v=$the universe element of 5.10; and by 5.3, $z'=v$.

5.12. $a+z=a$.

By 5.11 and 5.4, $a+z=a+(a+a')'=a$.

5.13. $a+v=v$.

By 5.10, 5.12, 5.7, 5.5, 5.3, 5.11, 5.12, 5.3, $v=a+a'=a+(a+z)'=a+(v'+a)'=[(v''+a)'+(a'+a)']'=[(v+a)'+v']'=[(a+v)'+z]'=(a+v)''=a+v$.

5.14. Definition. $ab=(a'+b')'$.

By 5.1 and 5.2, if a and b are in the class K, then ab is in the class K.

5.15. $aa=a$.

By 5.14, 5.6, 5.3, $aa=(a'+a')'=(a')'=a$.

5.16. $ab=ba$. (By 5.14 and 5.7.)

5.17. $aa'=z$.

By 5.14, 5.10, 5.11, $aa'=(a'+a)'=v'=z$.

5.18. $az=z$.

By 5.14, 5.11, 5.13, 5.11, $az=(a'+z')'=(a'+v)'=v'=z$.

5.19. $av=a$.

By 5.14, 5.11, 5.12, 5.3, $av=(a'+v')'=(a'+z)'=(a')'=a$.

5.20. $a+b=(a'b')'$. (By 5.14 and 5.3.)

5.21. $a+bc=(a+b)(a+c)$.

By 5.14, 5.5 and 5.3, 5.14, 5.7, $a+bc=a+(b'+c')'=[(b+a)'+(c+a)']'$ $=(b+a)(c+a)=(a+b)(a+c)$.

5.22. $a(b+c)=ab+ac$.

By 5.14, 5.14 and 5.3, 5.21, 5.14 and 5.3, 5.14, we have $a(b+c)=$ $[a'+(b+c)']'=[a'+b'c']'=[(a'+b')(a'+c')]'=(a'+b')'+(a'+c')'=ab+ac$.

5.23. $a+ab=a$.

By 5.19, 5.22, 5.13 and 5.7, 5.19, $a+ab=av+ab=a(v+b)=av=a$.

5.24. $a(a+b)=a$.

By 5.22, 5.15, 5.23, $a(a+b)=aa+ab=a+ab=a$.

5.25. *If $a'+b=v$ and $b'+a=v$, then $a=b$.*

By 5.11 and 5.12, $a+v'=a$. By 5.8, $(a'+b')'+(a'+b)'=a$. Hence if $a'+b=v$, then by 5.12, $a'+b'=a$. By 5.8, $(b'+a')'+(b'+a)'=b$. Hence if $b'+a=v$, then by 5.12, $b'+a'=b$. Hence by 5.7, $a=b$.

5.26. *If $a+b=v$ and $ab=z$, then $a'=b$.*

From $a+b=v$, by 5.3, $a''+b=v$. From $ab=z$, by 5.14, $(a'+b')'=z$, whence by 5.3, 5.7, 5.11, $b'+a'=v$. Hence by 5.25, $a'=b$.

5.27. $(a+b)+c=a+(b+c)$.

We prove first the following lemmas.

LEMMA (1). *If $x=(a+b)+c$, then $ax=a$, $bx=b$, and $cx=c$.*

For, by 5.7, 5.22, 5.24, 5.23,

$$ax = a[(a + b) + c] = a(a + b) + ac = a + ac = a,$$

whence similarly, $bx=b$; and by 5.7, 5.22, 5.15, 5.19, 5.22, 5.13, 5.19, $cx=c[(a+b)+c]=c(a+b)+cc=c(a+b)+c=c(a+b)+cv=c[(a+b)+v]=cv$ $=c$.

LEMMA (2). *If $x=(a+b)+c$, then $x'a=z$, $x'b=z$, and $x'c=z$.*

For, by 5.19 and 5.7, 5.10, 5.21, Lemma (1), 5.10, $x+a'=v(a'+x)$ $=(a'+a)(a'+x)=a'+ax=a'+a=v$, whence by 5.14, 5.11, $x'a=(x+a')'$ $=v'=z$. Similarly, $x'b=z$ and $x'c=z$.

LEMMA (3). *If $y=a+(b+c)$, then $y+a'=v$, $y+b'=v$, and $y+c'=v$.*

For, by Lemma (2), with 5.7, $y'a=z$, $y'b=z$, $y'c=z$, whence, by 5.20, 5.11, $y+a'=(y'a)'=z'=v$. Similarly $y+b'=v$ and $y+c'=v$.

The proof of the main theorem then proceeds as follows. By 5.22, 5.22, 5.6, $x'y=x'[a+(b+c)]=x'a+x'(b+c)=x'a+(x'b+x'c)=z+(z+z)=z+z=z$; and by 5.7, 5.20 and 5.3, 5.21, 5.21, 5.19,

$$x'+y=y+[c+(b+a)]'=y+[c'(b+a)']=y+[c'(b'a')]$$
$$=(y+c')[y+(b'a')]=(y+c')[(y+b')(y+a')]=v(vv)=vv=v.$$

Hence by 5.26, $x=y$; that is, $(a+b)+c=a+(b+c)$.

We can now establish the equivalence of the fifth set and the fourth set, as follows:

Theorems 5.1, 5.2, 5.6, 5.7, 5.8, and 5.27 show that all the postulates of the fourth set are deducible from Postulates 5.1–5.5; and conversely all the postulates of the fifth set are readily deducible from Postulates 4.1–4.6.

Incidentally, Theorems 5.1, 5.14, 5.12, 5.19, 5.7, 5.16, 5.21, 5.22, 5.10 and 5.17 show directly that all the postulates of the "first set" are deducible from Postulates 5.1–5.5 (when the product ab is defined as in 5.14); and conversely, all the postulates of the fifth set are readily deducible from Postulates Ia–VI (when a' is defined in terms of $+$ and $\times$ as in V); so that the equivalence between the fifth set and the "first set" is established directly, without reference to the fourth set.

To show that the fifth set of postulates is equivalent to Sheffer's set of 1913, which occupies so important a position in the revised edition of the *Principia Mathematica* (volume 1, 1925), we need the definition of Sheffer's "stroke" function, namely:

5.28. Definition. $a|b=(a+b)'=$ the "reject" of a and b (pronounced a per b).

On the basis of this definition we deduce Sheffer's postulates from Postulates 5.1–5.5 as follows:

(1) There are at least two distinct K-elements.

(2) Whenever a and b are K-elements, $a|b$ is a K-element. (By 5.1, 5.2.)

Definition. $a'=a|a$. (By 5.6.)

(3) $(a')'=a$. (By 5.3.)

(4) $a|(b|b')=a'$. (By 5.4 and 5.3.)

(5) $[a|(b|c)]'=(b'|a)|(c'|a)$. (By 5.5 and 5.3.)*

---

* In a paper published in 1916, B. A. Bernstein showed that Sheffer's postulates (3), (4), and (5) may be replaced by two postulates, P_3 and P_4:

P_3. $(b|a)|(b'|a)=a$.

P_4. $a'|(b'|c)=[(b|a')|(c'|a')]'$.

This change does not lead to any corresponding reduction in Postulates 5.1–5.5.

Independence proofs for the fifth set

The *independence* of the postulates of the fifth set is established by the existence of the following examples of systems $(K, +, ')$, each of which violates the like-numbered postulate, and satisfies all the other postulates of the set.

Example 5.1. K = two elements, 0, 1, with $a+b$ and a' defined as in the accompanying table, where x is any object not an element of the class K.

+	0	1	′
0	0	1	1
1	1	x	0

Example 5.2. K = two elements, 0, 1, with $a+b$ and a' given by the table (x being any object not an element of the class K).

+	0	1	′
0	0	1	x
1	1	1	x

Example 5.3. K = six elements, with $a+b$ and a' given by the table.

+	0	1	2	3	4	5	′
0	0	1	4	5	2	3	1
1	1	1	1	1	1	1	0
2	2	1	0	1	2	1	3
3	3	1	1	0	1	3	4
4	4	1	4	1	0	1	5
5	5	1	1	5	1	0	2

Here Postulate 5.3 fails, since $(2')'=3'=4$. All the other postulates of the fifth set are found to be satisfied.

It is interesting to note that while the commutative law, $a+b=b+a$, does not hold in this example, it is always true that $a+b=(b+a)''$.

Example 5.4. K = three elements, 0, 1, 2, with $a+b$ and a' given by the table.

+	0	1	2	′
0	0	1	2	1
1	1	1	1	0
2	2	1	2	2

Here Postulate 5.4 fails, since $0+(2+2')'=2$.

Example 5.5. K = three elements, 0, 1, 2, with $a+b$ and a' given by the table.

$+$	0	1	2	$'$
0	0	2	1	0
1	1	1	0	2
2	2	0	2	1

To show that Postulate 5.5 fails, take $a=1, b=1, c=2$.

The sixth set

The following set of postulates for Boolean algebra in terms of $(K, +, ')$ is suggested by B. A. Bernstein's version of the primitive propositions of the *Principia*. The only modifications are as follows: (a) his proposition 1.5 is omitted because it can be proved as a theorem; (b) his notation "$=1$" is here replaced by the notation "is in a subclass T", which corresponds more nearly to the Frege assertion sign, $\vdash$; and (c) our Postulate 6.8 is an additional postulate, not included among the primitive propositions of the *Principia*.

A trivial preliminary postulate, 6.0, demanding that the class K contain at least two distinct elements, is assumed without further mention; and in Postulates 6.4–6.8 the indicated combinations are assumed to be elements of K. Also, Postulates A, B, C, D are assumed without further mention. The numbers in brackets indicate the corresponding postulates in the Bernstein-*Principia* list.

Postulate 6.1. [1.71.] *If a and b are in the class K, then $a+b$ is in the class K.*

Postulate 6.2. [1.7.] *If a is in the class K, then a' is in the class K.*

There exists in the class K a *subclass* T having the following five properties:

Postulate 6.3. [1.1.] *If a is in T and $a'+b$ is in T, then b is in T.*

Postulate 6.4. [1.2.] *If a is in K, then $(a+a)'+a$ is in T.*

Postulate 6.5. [1.3.] *If a and b are in K, then $b'+(a+b)$ is in T.*

Postulate 6.6. [1.4.] *If a and b are in K, then $(a+b)'+(b+a)$ is in T.*

Postulate 6.7. [1.6.] *If a, b, c are in K, then $(b'+c)'+[(a+b)'+(a+c)]$ is in T.*

Postulate 6.8. *If T is a subclass having the five properties just mentioned, then we have: If $a'+b$ is in T, and $b'+a$ is in T, then $a=b$.*

The *consistency* of these postulates is established by the existence of any Boolean algebra $(K, +, ')$, with the subclass T taken as the class containing the single element v.

The *equivalence* of the sixth set and the fourth set is established as follows.

In the first place, all the postulates of the sixth set are readily deducible from the fourth set; the single element v (see 4.12) constitutes the required subclass T.

We proceed to show, conversely, that all the postulates of the fourth set can be derived as theorems from the sixth set.

6.9. *If $a'+b$ is in T and $b'+c$ is in T, then $a'+c$ is in T.*

By 6.7, $(b'+c)'+[(a'+b)'+(a'+c)]$ is in T. But by hypothesis, $b'+c$ is in T. Hence by 6.3, $(a'+b)'+(a'+c)$ is in T. But by hypothesis, $a'+b$ is in T. Hence by 6.3, $a'+c$ is in T.

Note. This theorem 6.9 corresponds to the "syllogism" in the theory of deduction, while 6.3 corresponds to the "rule of inference."

By the aid of this theorem we can establish at once the redundancy of proposition 1.5 in the Bernstein-*Principia* list. This theorem 1.5 will serve as a lemma in the proof of the associative law (6.12).

6.10. [1.5.] $[a+(b+c)]'+[b+(a+c)]$ *is in T.*

(The following proof is adapted from a proof given, in another notation, by P. Bernays in 1926. It does not involve Postulate 6.8.)

By 6.5, $c'+(a+c)$ is in T.

By 6.7, $[c'+(a+c)]'+\{(b+c)'+[b+(a+c)]\}$ is in T.

Hence by 6.3, $(b+c)'+[b+(a+c)]$ is in T.

By 6.7, $\{(b+c)'+[b+(a+c)]\}'+([a+(b+c)]'+\{a+[b+(a+c)]\})$ is in T.

Hence by 6.3, $[a+(b+c)]'+\{a+[b+(a+c)]\}$ is in T.

By 6.6, $\{a+[b+(a+c)]\}'+\{[b+(a+c)]+a\}$ is in T.

Hence by 6.9,

$$[a+(b+c)]'+\{[b+(a+c)]+a\} \text{ is in } T. \tag{1}$$

By 6.5, $a'+(c+a)$ is in T.

By 6.6, $(c+a)'+(a+c)$ is in T.

Hence by 6.9, $a'+(a+c)$ is in T.

By 6.5, $(a+c)'+[b+(a+c)]$ is in T.

Hence by 6.9, $a'+[b+(a+c)]$ is in T.

By 6.7, $\{a'+[b+(a+c)]\}'+(\{[b+(a+c)]+a\}'+\{[b+(a+c)]+[b+(a+c)]\})$ is in T.

Hence by 6.3, $\{[b+(a+c)]+a\}'+\{[b+(a+c)]+[b+(a+c)]\}$ is in T.

By 6.4, $\{[b+(a+c)]+[b+(a+c)]\}'+[b+(a+c)]$ is in T.

Hence by 6.9,

$$\{[b+(a+c)]+a\}'+[b+(a+c)] \text{ is in } T. \tag{2}$$

From (1) and (2), by 6.9, $[a+(b+c)]'+[b+(a+c)]$ is in T.

The next three theorems correspond to Postulates 4.3, 4.4, 4.5 of the fourth set.

6.11. $a+b=b+a$.

By 6.6, $(a+b)'+(b+a)$ is in T. Again, by 6.6, $(b+a)'+(a+b)$ is in T. Hence by 6.8, $a+b=b+a$.

6.12. $(a+b)+c=a+(b+c)$.

By 6.10, $[a+(c+b)]'+[c+(a+b)]$ is in T. Hence by 6.11,

(1) $$[a + (b + c)]' + [(a + b) + c] \text{ is in } T.$$

From (1), $[c+(b+a)]'+[(c+b)+a]$ is in T. Hence by 6.11,

(2) $$[(a + b) + c]' + [a + (b + c)] \text{ is in } T.$$

From (1) and (2), by 6.8, $a+(b+c)=(a+b)+c$.

6.13. $a+a=a$.

By 6.4, $(a+a)'+a$ is in T. By 6.5, $a'+(a+a)$ is in T. Hence by 6.8, $a+a=a$.

The following theorems establish the existence and properties of the universe element v.

6.14. *If a is in K, then $a+a'$ is in T.*

By 6.7 $[(a+a)'+a]'+\{[a'+(a+a)]'+(a'+a)\}$ is in T. But by 6.4, $(a+a)'+a$ is in T. Hence by 6.3, $[a'+(a+a)]'+(a'+a)$ is in T. But by 6.5, $a'+(a+a)$ is in T. Hence by 6.3, $a'+a$ is in T. Hence by 6.11, $a+a'$ is in T.

6.15. $a+a'=b+b'$.

By 6.7, $[(a+b)'+(b+a)]'+\{[b'+(a+b)]'+[b'+(b+a)]\}$ is in T. But by 6.6, $(a+b)'+(b+a)$ is in T. Hence by 6.3, $[b'+(a+b)]'+[b'+(b+a)]$ is in T. But by 6.5, $b'+(a+b)$ is in T. Hence by 6.3,

(1) $$b' + (b + a) \text{ is in } T.$$

From (1), $(a+a')'+[(a+a')+(b+b')']$ is in T. But by 6.14, $a+a'$ is in T. Hence by 6.3,

(2) $$(a + a') + (b + b')' \text{ is in } T.$$

Now by 6.6, $[(a+a')+(b+b')']'+[(b+b')'+(a+a')]$ is in T. Hence by (2) and 6.3,

(3) $$(b + b')' + (a + a') \text{ is in } T.$$

From (3),

(4) $$(a + a')' + (b + b') \text{ is in } T.$$

From (4) and (3), by 6.8, $b+b'=a+a'$.

This theorem justifies the following definition.

6.16. Definition. $v=a+a'=$the "universe" element of the system.

This element v exists, by 6.1 and 6.2, and is unique, by 6.15. Also, by 6.11, $v=a'+a$; and obviously, by 6.14, v is in T.

6.17. *If a is in T, then $a=v$.*

By 6.5,

(1) $$a' + (a' + a) \text{ is in } T.$$

Again, by 6.5, $a'+[(a'+a)'+a]$ is in T. But by hypothesis, a is in T. Hence by 6.3,

(2) $$(a' + a)' + a \text{ is in } T.$$

From (1) and (2), by 6.8, $a=a'+a$. Hence by 6.16, $a=v$.

This theorem shows that the subclass T consists of a single element of the class K, namely, the element v defined in 6.16.

6.18. $a+v=v$ (*whence, by* 6.11, $v+a=v$).

By 6.5, $v'+(a+v)$ is in T. But by 6.16, v is in T. Hence by 6.3, $a+v$ is in T. Hence by 6.17, $a+v=v$.

6.19. $v'+a=a$ (*whence, by* 6.11, $a+v'=a$).

By 6.7, $(v'+a)'+[(a+v)'+(a+a)]$ is in T. Hence by 6.18 and 6.13, $(v'+a)'+(v'+a)$ is in T. Hence by 6.11 and 6.12, $v'+[(v'+a)'+a]$ is in T. Hence by 6.3, $(v'+a)'+a$ is in T. But by 6.5, $a'+(v'+a)$ is in T. Hence by 6.8, $v'+a=a$.

6.20. $v'\neq v$.

Suppose $v'=v$. Then by 6.19 we would have $a+v=a$. But by 6.18, $a+v=v$. Hence we would have $a=v$; that is, every element a would be equal to the element v, so that the class K would contain only a single element. This trivial case is excluded by Postulate 6.0.

6.21. *$a'=a$ is always false.*

Suppose there existed an element c such that $c'=c$. Then by 6.13 and 6.16, $c=c+c=c+c'=v$. Hence $c'=v'$. Hence, if $c'=c$ we would have $v'=v$, which by 6.20 is impossible.

6.22. *If a is in T, then a' is not in T.*

Since a is in T, by 6.17, $a=v$, whence by 6.2, $a'=v'$. If a' also were in T, then, by 6.17, $a'=v$, whence $v'=v$, which by 6.20 is impossible.

6.23. $a''=a$.

By 6.7, $(a''+a''')'+[(a+a')'+(a+a''')]$ is in T. But by 6.14, $a''+a'''$ is in T. Hence by 6.3, $(a+a')'+(a+a''')$ is in T. But by 6.14, $a+a'$ is in T.

Hence by 6.3, $a+a'''$ is in T, whence by 6.11,

(1) $$a''' + a \text{ is in } T.$$

Again, by 6.14,

(2) $$a' + a'' \text{ is in } T.$$

From (1) and (2), by 6.8, $a''=a$.

6.24. Definition. $ab=(a'+b')'$.

6.25. $ab=ba$. (By 6.24 and 6.11.)

6.26. $(ab)c=a(bc)$. (By 6.24, 6.23, 6.12.)

6.27. $aa=a$. (By 6.24, 6.23, 6.13.)

6.28. $av=a$.

By 6.24, 6.11, 6.19, 6.23, $av=(a'+v')'=(v'+a')'=(a')'=a$.

6.29. Definition. $z=v'=$the "zero" element of the system.

6.30. $a+z=a$. (By 6.29 and 6.19.)

6.31. $az=z$.

By 6.24, 6.29, 6.23, 6.18, $az=(a'+z')'=(a'+v)'=v'=z$.

6.32. $aa'=z$.

By 6.24, 6.16, 6.29, $aa'=(a'+a'')'=v'=z$.

6.33. $(ab)'=a'+b'$. (By 6.24, 6.23.)

6.34. $(a+b)'=a'b'$. (By 6.24, 6.23.)

In the following proofs, tacit use will be made of 6.17.

6.35. *If $a'+b=v$, then $ab=a$.*

By 6.7, $[b'+(a'+ab)]'+\{(a'+b)'+[a'+(a'+ab)]\}=v$. But by 6.11, 6.12, 6.24, 6.16, $b'+(a'+ab)=(a'+b')+ab=(a'+b')+(a'+b')'=v$. Hence by 6.3, $(a'+b)'+[a'+(a'+ab)]=v$. But by hypothesis, $a'+b=v$. Hence by 6.3, $a'+(a'+ab)=v$, whence by 6.12 and 6.13,

(1) $$a' + ab = v.$$

Again, by 6.33, 6.11, 6.12, 6.16, $(ab)'+a=(a'+b')+a=b'+(a+a')=b'+v$, whence by 6.18,

(2) $$(ab)' + a = v.$$

From (1) and (2), by 6.8, $a=ab$.

6.36. *If $ab'=z$, then $a+b=b$.*

From $ab'=z$, by 6.33, 6.25, 6.29, $b''+a'=(b'a)'=(ab')'=z'=v$. Hence by 6.35, $b'a'=b'$, whence by 6.34, $(b+a)'=b'$. Hence by 6.11, 6.23, $a+b=b$.

6.37. *If $ab'=z$, then $ab=a$.*

From $ab'=z$, by 6.24, 6.23, 6.29, $a'+b=v$. Hence by 6.35, $ab=a$.

6.38. *If $ab=a$, then $a+b=b$.*

From $ab=a$, by 6.33, $a'=a'+b'$. Hence by 6.12, 6.13, 6.16, 6.18, $(a')+b=(a'+b')+b=a'+(b+b')=a'+v=v$, whence by 6.25, 6.23, $b''+a'=v$. Hence by 6.35, $b'a'=b'$, whence by 6.34, $(b+a)'=b'$. Hence by 6.25, 6.23, $a+b=b$.

6.39. $a(ab+ab')'=z$.

Let $x=a(ab+ab')'$. Then by 6.26, 6.25, 6.32, 6.31,

$$(1)\quad x(ab+ab')=[a(ab+ab')'](ab+ab')=a[(ab+ab')'(ab+ab')]=az=z;$$

$$(2)\quad xa'=[a(ab+ab')']a'=(ab+ab')'(aa')=(ab+ab')z=z.$$

From (1), by 6.25, 6.23, $(ab+ab')x''=z$, whence by 6.36, $ab+ab'+x'=x'$. Hence by 6.11, 6.12, 6.13, $ab+x'=ab+(ab+ab'+x')=ab+ab'+x'=x'$, whence by 6.24, 6.23,

$$(3)\quad (ab)'x=x;$$

also, $ab'+x'=ab'+(ab+ab'+x')=ab+ab'+x'=x'$, whence by 6.24, 6.23,

$$(4)\quad (ab')'x=x.$$

From (4), by 6.26, 6.25, 6.32, 6.31,

$$(xa)b'=x(ab')=[(ab')'x](ab')=x[(ab')(ab')']=xz=z,$$

whence by 6.37, $(xa)b=xa$.

From (2), $xa'=z$, whence by 6.37, $xa=x$. Hence $(xa)b=x$, whence by 6.26,

$$(5)\quad x(ab)=x.$$

From (3) and (5), by 6.27, 6.26, 6.25, 6.32, 6.31,

$$x=xx=[x(ab)][x(ab)']=x[(ab)(ab)']=xz=z.$$

6.40. $ab+ab'=a$.

By 6.25, 6.26, 6.27, $(ab)a=a(ab)=(aa)b=ab$, whence by 6.38,

$$(1)\quad ab+a=a.$$

From (1),

$$(2)\quad ab'+a=a.$$

Hence $(ab+a)+(ab'+a)=a+a$, whence by 6.25, 6.26, 6.27, $a+(ab+ab')=a$. But by 6.39, $a(ab+ab')'=z$, whence by 6.36, $a+(ab+ab')=ab+ab'$. Therefore $ab+ab'=a$.

6.41. $(a'+b')'+(a'+b)'=a$. (From 6.40, by 6.24 and 6.23.)

The proof of the equivalence of the sixth set and the fourth set is thus complete; Theorems 6.1, 6.2, 6.11, 6.12, 6.13, and 6.41 show that all the postulates of the fourth set are deducible from the sixth set.

INDEPENDENCE PROOFS FOR THE SIXTH SET

We first give three examples for the independence of Postulate 6.8.

Example 6.8 (1). K = three elements, 0, 1, 2; with $a+b$ and a' given by the table.

+	0	1	2	′
0	0	1	2	1
1	1	1	1	0
2	2	1	2	1

This system (K, $+$, $'$) has all the properties called for by Postulates 6.1–6.7, with the subclass T consisting of the single element 1. The system fails on 6.8, since $2'+0=1$ and $0'+2=1$, but not $2=0$. It is interesting to note that $a'''+a$ in T and $a'+a''$ in T both hold, but $a''=a$ is not true when $a=2$. We note also that $a+b=b+a$ and $(a+b)+c=a+(b+c)$ and $a+a=a$; further, if $a+b$ is in T then a is in T or b is in T.

Example 6.8 (2). K = five elements. (This example was suggested to me, in another connection, by Dr. K. E. Rosinger.)

+	0	1	2	3	4	′
0	0	1	2	3	4	1
1	1	1	1	1	1	0
2	2	1	2	3	1	4
3	3	1	3	3	1	4
4	4	1	1	1	4	3

Postulates 6.1–6.7 hold, with the subclass T consisting of the single element 1. Postulate 6.8 fails, since $2'+3=1$ and $3'+2=1$, but not $2=3$. Here $a+b=b+a$ and $(a+b)+c=a+(b+c)$ and $a+a=a$ are always true, but not $a''=a$. Further, $a+b$ can equal 1 when neither $a=1$ nor $b=1$.

Example 6.8 (3). K = six elements. (This example was suggested to me by Mr. P. Henle.)

+	0	1	2	3	4	5	′
0	0	1	2	3	4	5	1
1	1	1	1	1	1	1	0
2	2	1	5	1	1	5	3
3	3	1	1	3	4	1	2
4	4	1	1	3	4	1	5
5	5	1	2	1	1	2	4

Postulates 6.1–6.7 will be found to hold, with the subclass T consisting of the single element 1. Postulate 6.8 fails, since $2'+5=1$ and $5'+2=1$, but not $2=5$. We note that $a''=a$ holds. Also, $(a+b)'+(b+a)=1$, but not $a+b=b+a$. Also, $(a+a)'+a=1$ and $a'+(a+a)=1$, but not $a+a=a$.

Obvious examples for Postulate 6.1 and 6.2 are the following, in which x is any object not in the class K.

Example 6.1

+	0	1	2	3	′
0	0	1	2	3	1
1	1	1	1	1	0
2	2	1	2	x	3
3	3	1	x	3	2

Example 6.2

+	0	1	2	3	′
0	0	1	2	3	1
1	1	1	1	1	0
2	2	1	2	1	x
3	3	1	1	3	x

The remaining examples (for 6.3–6.7) I take from a recent paper by P. Henle (*The independence of the postulates of logic*, Bulletin of the American Mathematical Society, 1932).

Example 6.3

+	0	1	′
0	1	1	1
1	1	1	0

Example 6.4

+	0	1	2	′
0	0	1	2	1
1	1	1	1	0
2	2	1	1	2

Example 6.5

+	0	1	2	3	′
0	0	1	0	0	1
1	1	1	1	1	0
2	0	1	0	1	3
3	0	1	1	0	2

Example 6.6

+	0	1	2	3	′
0	0	1	2	3	1
1	1	1	1	1	0
2	2	1	2	1	3
3	0	1	1	3	2

Example 6.7

+	0	1	2	3	4	5	′
0	0	1	2	3	4	5	1
1	1	1	1	1	1	1	0
2	2	1	2	1	1	1	3
3	3	1	1	3	4	5	2
4	4	1	1	4	4	1	5
5	5	1	1	5	1	5	4

The following unsolved problem may be noted. If 6.8 were replaced by the same postulate without the qualifying clause (call it 6.8a), then the independence of 6.3 would become an open question (since the present Example 6.3 does not satisfy 6.8a).

Appendix I

The connection between Boolean algebra and the Principia

In order to establish the connection between Boolean algebra and the system set forth in Section A of the *Principia*, we first quote the following propositions verbatim from the second edition of the *Principia*.

*1.71. If p and q are elementary propositions, $p \vee q$ is an elementary proposition.

*1.7. If p is an elementary proposition, $\sim p$ is an elementary proposition.

*4.31. $\vdash : p \vee q \cdot \equiv \cdot q \vee p$.

*4.33. $\vdash : (p \vee q) \vee r \cdot \equiv \cdot p \vee (q \vee r)$.

*4.25. $\vdash : p \cdot \equiv \cdot p \vee p$.

*4.5. $\vdash : p \cdot q \cdot \equiv \cdot \sim(\sim p \vee \sim q)$.

*4.42. $\vdash :\cdot p \cdot \equiv : p \cdot q \cdot \vee \cdot p \cdot \sim q$.

If now we call the class of "elementary propositions" the class K, and write $p+q$ for $p \vee q$, and p' for $\sim p$, these propositions become the following:

*1.71. If p and q are in the class K, then $p+q$ is in the class K.

*1.7. If p is in the class K, then p' is in the class K.

*4.31. $p+q \equiv q+p$.

*4.33. $(p+q)+r \equiv p+(q+r)$.

*4.25. $p \equiv p+p$.

*4.5. $pq \equiv (p'+q')'$.

*4.42. $p \equiv pq+pq'$.

But these propositions are precisely the same as the postulates of our fourth set for Boolean algebra, except that the sign $\equiv$ occurs in place of the sign $=$.

It remains, therefore, to examine the properties of the sign $\equiv$ as used in the *Principia*, in comparison with Postulates A, B, C, D governing the use of the sign $=$.

Here it is necessary to distinguish between the formal statements and the informal statements in the *Principia*. Among the formal statements we find

*4.2. $\vdash . p \equiv p$,

which is the same as Postulate A.

Another formal statement (in view of *4.01 and *1.01) is

*4.21. $\vdash : (p \equiv q)' \cdot \vee \cdot (q \equiv p)$.

In accordance with the informal statement under (6) in *1, this formula means

"$p \equiv q$ is false or $q \equiv p$ is true,"

and this in turn means

"If $p \equiv q$, then $q \equiv p$,"

which is the same as Postulate B.

Among the informal statements we find under *4.22 that "$\equiv$" denotes a *relation*, namely the "relation of equivalence," and that "the relation of equivalence is reflexive (*4.2), symmetrical (*4.21) and transitive (*4.22)."†

We have already cited *4.2 and *4.21. In regard to *4.22, the formal statement of this theorem (in view of *1.01) is

*4.22. $\vdash : \sim(p \equiv q \cdot q \equiv r) \cdot \vee \cdot (p \equiv r)$;

and in accordance with the informal statement under (6) in *1, this formula means "$p \equiv q \cdot q \equiv r$ is false or $p \equiv r$ is true." This in turn means

"If $p \equiv q$ and $q \equiv r$, then $p \equiv r$,"

which is the same as Postulate C.

Again, among the informal statements we find under *4.01 the following:

"If $p \equiv q$, then q may be substituted for p without altering the truth-value of any function of p which involves no primitive ideas except those enumerated in *1."

This comes to the same thing as our Postulate D.

Hence, if we accept the above mentioned informal statements as a valid part of the theory of the *Principia*, we have the following theorem:

THEOREM I. *With respect to* $(K, \vee, \sim, \equiv)$, *the informal system of the Principia is a Boolean algebra.*

Here, from the abstract postulational point of view, K is an undefined class; $\vee$ (or $+$) is an undefined binary operator; $\sim$(or $'$) is an undefined unary operator; and $\equiv$ is an undefined relation. Concretely, K may be interpreted as the class of "elementary functions"; $a \vee b$ as "a or b"; $\sim a$ as "not-a"; and $a \equiv b$ as "a equivalent to b." But any other interpretation of the symbols $(K, \vee, \sim, \equiv)$ which satisfies the rules laid down would be a valid example of the system.

One further question arises. Since the number of elements in a Boolean algebra may be any power of 2, it is interesting to inquire whether there is anything in the *Principia* which restricts the number of elements.

Among the formal statements we find

*5.15. $\vdash : p \equiv q \cdot \vee \cdot p \equiv \sim q$;

and according to the informal statement under (6) in *1, this formula means

"Either $p \equiv q$ is true or $p \equiv \sim q$ is true."

That is, if q is any particular element of the class K, then every other element must be equivalent either to q or to q'; so that there are only two non-

† In the formal part of the *Principia* there is a different definition of $\equiv$, which does not concern us here.

equivalent elements in the system. Hence, if we accept the informal as well as the formal statements, we have

THEOREM II. *With respect to* $(K, \vee, \sim, \equiv)$, *the informal system of the Principia is a Boolean algebra containing only two non-equivalent elements.*

APPENDIX II

A SET OF INDEPENDENT POSTULATES FOR PRINCIPIA MATHEMATICA

This appendix contains a revision of the primitive propositions of the *Principia* (Section A), without making use of the equality sign, or the Boolean notation "$=1$." (Compare Bernstein's papers of 1931 and 1932.)

The *primitive ideas* in this theory are four in number:

K = an undefined class of elements, $a, b, c, \cdots$ (K being interpretable as the class of "propositions");

T = an undefined subclass in K (T being interpretable as the subclass of "true" propositions, indicated in the *Principia* by the assertion sign $\vdash$);

$a+b$ = the result of an undefined binary operation ($a+b$ being interpretable as "a or b," denoted in the *Principia* by $a \vee b$);

a' = the result of an undefined unary operation (a' being interpretable as "not-a," denoted in the *Principia* by $\sim a$).

The *postulates* here assumed are seven in number, the first five corresponding precisely to "formal," and the last two to "informal" statements in the *Principia:*

POSTULATE 1. *If a and b are in K, then $a+b$ is in K.* [*1.71]

POSTULATE 2. *If a is in K, then a' is in K.* [*1.7]

POSTULATE 3. *If a, b, etc. are in K, then $b'+(a+b)$ is in T.* [*1.3]

POSTULATE 4. *If a, b, etc. are in K, then $(a+b)'+(b+a)$ is in T.* [*1.4]

POSTULATE 5. *If a, b, c, etc. are in K, then $(b'+c)'+[(a+b)'+(a+c)]$ is in T.* [*1.6]

POSTULATE 6. *If $a+b$ is in T, then at least one of the elements a and b is in T.*

POSTULATE 7. *If a' is in T, then a is not in T.*

From these postulates the following propositions are deducible as theorems.†

† The proof of 8 follows at once from 6.

The proof of 9 depends on the following lemmas:

(a) If $a+b$ is in T, then $b+a$ is in T. (By 4, 6, 7.)

(b) If b is in T, then $a+b$ is in T. (By 3, 6, 7.)

(c) If a is in T, then $a+b$ is in T. (By (a) and (b).)

(d) If a is not in T and b is not in T, then $a+b$ is not in T. (By 6.)

(e) If $a+b$ is not in T, then a is not in T and b is not in T. (By (b) and (c).)

If 9 were false, we should have, by (e) and (d), $(a+a)'+(a+a)$ not in T, contrary to 4. The proof of 10 (due to Bernays in 1926) is given in 6.10 above.

8. *If a is in T and $a'+b$ is in T, then b is in T.* [*1.1]

9. *If a, $a+a$, etc. are in K, then $(a+a)'+a$ is in T.* [*1.2]

10. *If a, b, c, etc. are in K, then $[a+(b+c)]'+[b+(a+c)]$ is in T.* [*1.5]

Any system $(K, T, +, ')$ which satisfies Postulates 1–7 may be called an "*informal Principia system*," since all the propositions, both "formal" and "informal," in Section A of the Principia are deducible from these seven postulates.

In regard to the definitions of $a \supset b$ and $a \equiv b$, the "formal" and "informal" statements in the *Principia* are not in precise agreement. In the "formal" part (see *1.01 and *4.01), $a \supset b$ and $a \equiv b$ are defined as *elements* determined by a and b (analogous to $a+b$). But in the "informal" part (and in practically all common usage) the symbols $\supset$ and $\equiv$ are used to indicate *relations* between the two elements. These two usages may be reconciled by defining the symbols $a \supset b$ and $a \equiv b$ as elements, and the words "a implies b" and "a is equivalent to b" as statements about these elements.

11a. **Definition.** $a \supset b$ means $a'+b$. [*1.01]

11b. **Definition.** (a implies b) means ($a'+b$ is in T).

That is, "a implies b" means that the element $a \supset b$ is in T.

12a. **Definition.** $a \equiv b$ means $[(a'+b)'+(b'+a)']'$. [*4.01]

12b. **Definition.** (a is equivalent to b) means $\{[(a'+b)'+(b'+a)']'$ *is in* $T\}$.

That is, "a is equivalent to b" means that the element $a \equiv b$ is in T.

From Postulates 1–7, with the aid of these definitions, the following much-discussed theorems are deducible:

13. *If a is in K, and b is in T, then $a \supset b$ is in T.* [*2.02]

That is, "a true proposition is implied by any proposition."

14. *If a is not in T and b is in K, then $a \supset b$ is in T.* [*2.21]

That is, "a false proposition implies any proposition."

15. *If a is in T and b is in T, then $a \equiv b$ is in T.* [*5.1]

That is, "two propositions are equivalent if they are both true."

16. *If a is not in T and b is not in T, then $a \equiv b$ is in T.* [*5.21]

That is, "two propositions are equivalent if they are both false."

Hence an "informal *Principia* system," as above defined, contains only two "non-equivalent" elements.

17. *If $a'+(b+c)$ is in T, then $(a'+b)+(a'+c)$ is in T.* [*4.78]

18. *If $a'+(b+c)$ is in T, then at least one of the elements $a'+b$ and $a'+c$ is in T.*

That is, if a implies $b+c$, then a implies b or a implies c.

It can be shown, by definite examples of systems $(K, T, +, ')$, that Theorems 14, 16, and 18 cannot be deduced from the "formal" part of the

Principia (1–5, 8–10), without the aid of the "informal" statement here listed as Postulate 6. (See Example 6, below.)

The *consistency* of Postulates 1–7 is shown by any one of the three following examples of systems (K, T, $+$, $'$) in which all seven postulates are satisfied.

Example 0.1.

K = a class of four numbers, say 0, 1, 2, 3;
T = the class containing the two numbers 1 and 3;
$a+b$ = the number given by the table;
a' = the number given by the table.

+	0	1	2	3	′
0	0	1	2	3	1
1	1	1	1	1	0
2	2	1	2	1	3
3	3	1	1	3	2

Example 0.2.

K = a class of five numbers, say 2, 3, 4, 8, 9;
T = the class containing the three numbers 2, 3, 4;
$a+b$ and a' = the numbers given by the table.

+	2	3	4	8	9	′
2	2	2	3	4	2	8
3	2	2	4	4	2	8
4	3	4	3	4	2	9
8	4	4	4	8	9	2
9	3	3	3	9	9	3

(Here the selection and arrangement of elements in each of the four compartments of the table is arbitrary, provided the two groups 2, 3, 4 and 8, 9, are kept separate.)

Example 0.3.

K = a class of two numbers, 0, 1;
T = the class containing the single element 1;
$a+b$ and a' = the numbers given by the table.

+	0	1	′
0	0	1	1
1	1	1	0

The *independence* of Postulates 1–7 is shown by the following examples of systems $(K, T, +, ')$, each of which violates the like-numbered postulate and satisfies all the other six.

Example 1. Same as Example 6.1, with $T=1$.

Example 2. Same as Example 6.2, with $T=1, 3$.

Example 3. Same as Example 6.5, with $T=1, 3$.

Example 4. Same as Example 6.6, with $T=1, 3$.

Example 5. Same as Example 4.6, with $T=1, 2, 3, 4$.

Example 6. Same as Example 0.1, above, with $T=1$ instead of $T=1, 3$.

Example 7. Same as Example 0.1, above, with $T=0, 1, 2, 3$ instead of $T=1, 3$.

The last two examples satisfy not only Postulates 1–5, but also Theorems 8, 9, and 10; so that *Postulates* 6 *and* 7 *are not deducible from the "formal" part of the Principia*. This fact is of fundamental importance in any discussion of the adequacy of the "theory of deduction" as set forth in the formal part of the *Principia*.

HARVARD UNIVERSITY,
CAMBRIDGE, MASS.

SUFFICIENT CONDITIONS FOR A PROBLEM OF MAYER IN THE CALCULUS OF VARIATIONS*

BY

G. A. BLISS AND M. R. HESTENES

1. Introduction. The general problem of Mayer with variable end points as proposed by Bliss (V, p. 305)† is that of finding in a class of arcs

$$(1:1) \qquad y_i = y_i(x) \qquad (x_1 \leqq x \leqq x_2;\ i = 1, \cdots, n)$$

satisfying a system of differential equations and end conditions

$$\phi_\alpha(x, y, y') = 0 \qquad (\alpha = 1, \cdots, m < n),$$

$$\psi_\mu[x_1, y(x_1), x_2, y(x_2)] = 0 \qquad (\mu = 1, \cdots, p \leqq 2n + 1)$$

one which minimizes a function of the form

$$g[x_1, y(x_1), x_2, y(x_2)].$$

Bliss has shown that this problem is equivalent to a problem of Bolza (V, p. 306) in the sense that each can be transformed into one of the other type. For the problem of Bolza the function to be minimized is

$$I = g[x_1, y(x_1), x_2, y(x_2)] + \int_{x_1}^{x_2} f(x, y, y')dx,$$

and it is clear at once that the problem of Mayer is a problem of Bolza having $f \equiv 0$.

Sufficient conditions for the problem of Bolza have been established by Morse (XI, p. 528) and Bliss (XII, p. 271). However the hypotheses which they make, in particular that of normality on every sub-interval, imply that the function f is not identically zero, and the sets of sufficient conditions established by them are therefore not applicable to the problem of Mayer without further modification. In view of this fact it is the purpose of the authors of the present paper to establish a set of sufficient conditions for the problem of Mayer with variable end points. This will be done in two parts, the first of which is the paper here presented, dealing only with the special case in which the number of end conditions $\psi_\mu = 0$ is exactly $2n+1$. By methods similar to those used by Bliss for the problem of Bolza (XII, pp. 261–274) the results obtained will be extended to the general case in a second paper by Hestenes.

* Presented to the Society, April 8, 1932; received by the editors June 9, 1932.

† Roman numerals in parentheses refer to the bibliography at the end of this paper.

The problem considered here is an obvious generalization of the classical problem of Mayer and reduces to the latter when the expression to be minimized is the function $g = y_1(x_2)$ and the end conditions $\psi_\mu = 0$ are the conditions

$$x_1 - \alpha_1 = y_i(x_1) - \beta_{i1} = x_2 - \alpha_2 = y_j(x_2) - \beta_{j2} = 0$$
$$(i = 1, \cdots, n; j = 2, \cdots, n),$$

the α's and β's being constants. Sufficiency theorems for the classical problem have been established by Egorov (II, p. 376), Kneser (I, p. 250; VIII, p. 290), and Larew (VII, p. 65), who use in each case an n-dimensional field defined in the $(n+1)$-dimensional space of points $(x, y_1, \cdots, y_n)$ by an $(n-1)$-parameter family of extremals passing through a fixed point. Such a field does not seem to be applicable to the problem considered here, but one can use instead a field of $n+1$ dimensions defined by an n-parameter family of extremals in $(x, y_1, \cdots, y_n)$-space. The construction and use of such a field are important features of this paper. An $(n+1)$-dimensional field of this sort is applicable to the more special classical problem of Mayer also, and a fundamental sufficiency theorem for this case can be established in this way with greater ease and fewer restrictions than have hitherto been required.

2. **Preliminary remarks.** In the following pages it is assumed that the various indices have the following ranges unless otherwise explicitly specified:

$$i, k = 1, 2, \cdots, n; \qquad \alpha, \beta = 1, 2, \cdots, m < n;$$
$$\rho, \sigma = 1, 2, \cdots, 2n + 1; \qquad r = 1, 2, \cdots, n - 1;$$
$$s = 1, 2, \cdots, 2n - 1.$$

The tensor analysis summation convention is used freely throughout. We make the following hypotheses concerning a particular arc E_{12} whose minimizing properties are to be studied:

(a) The functions $y_i(x)$ defining E_{12} are continuous on the interval $x_1 x_2$, and this interval can be subdivided into a finite number of parts on each of which these functions have continuous derivatives.

(b) The functions ϕ_α have continuous partial derivatives of the first three orders in a neighborhood $\mathfrak{R}$ of the values (x, y, y') on E_{12}, and at each element (x, y, y') in $\mathfrak{R}$ the matrix $\|\phi_{\alpha y'_k}\|$ has rank m.

(c) The functions g, ψ_ρ have continuous partial derivatives of the first two orders in a neighborhood of the end values $(x_1, y_{i1}, x_2, y_{i2})$ of E_{12} in which the determinant

$$(2:1) \qquad \begin{vmatrix} g_{x_1} & g_{y_{i1}} & g_{x_2} & g_{y_{i2}} \\ \psi_{\rho x_1} & \psi_{\rho y_{i1}} & \psi_{\rho x_2} & \psi_{\rho y_{i2}} \end{vmatrix}$$

is different from zero.

An *admissible set* (x, y, y') is a set interior to $\mathfrak{R}$ and satisfying the equations $\phi_\alpha=0$. An arc (1:1) having the continuity properties described in (a) is called *admissible* if all of its elements (x, y, y') are admissible. The definitions of equations of variation and of admissible variations used in the following pages are those of Bliss (V, p. 307; IX, p. 677). The problem of Mayer here proposed can now be more precisely stated as that of finding in the class of admissible arcs satisfying the end conditions $\psi_\rho=0$ one which minimizes the function g.

I. THE FIRST NECESSARY CONDITION. *For every minimizing arc E_{12} for the problem of Mayer as here proposed there exist constants c_i and a function $F=\lambda_\alpha(x)\phi_\alpha$ such that the equations*

$$F_{y_i'} = \int_{x_1}^{x} F_{y_i}dx + c_i, \quad \phi_\alpha = 0 \tag{2:2}$$

are satisfied at each point of E_{12}. The multipliers $\lambda_\alpha(x)$ are continuous except possibly at the values of x defining corners of E_{12} and do not vanish simultaneously at any point of E_{12}.

To prove this theorem one needs only to combine the methods used by Bliss for the corresponding theorems in the problems of Mayer (V, p. 311) and Lagrange (IX, p. 683). It is also an immediate corollary of a theorem established by Morse and Myers for the problem of Bolza (X, p. 245).

THEOREM 2:1. *If the functions $\lambda_\alpha(x)$ are a set of multipliers with which an admissible arc E_{12} satisfies the equations* (2:2), *then for every set of admissible variations ξ_1, ξ_2, $\eta_i(x)$ along E_{12} the functions $\eta_i(x)$ satisfy the equations*

$$F_{y_i'}\,\eta_i \Big|_{x'}^{x''} = 0 \tag{2:3}$$

for every interval $x'x''$.

This result is readily provable by multiplying the equations of variation

$$\phi_{\alpha y_i}\eta_i + \phi_{\alpha y_i'}\eta_i' = 0$$

by the multipliers $\lambda_\alpha(x)$, adding, and applying the usual integration by parts with the help of equations (2:2).

An admissible arc E_{12} is said to be *normal relative to the end conditions* $\psi_\rho=0$ if there exist for it $2n+1$ sets of admissible variations ξ_1^σ, ξ_2^σ, $\eta_i^\sigma(x)$ such that the determinant $|\Psi_\rho(\xi^\sigma, \eta^\sigma)|$ is different from zero, where

$$\Psi_\rho(\xi, \eta) = (\psi_{\rho x_1} + y_{i1}'\,\psi_{\rho y_{i1}})\xi_1 + \psi_{\rho y_{i1}}\eta_{i1} + (\psi_{\rho x_2} + y_{i2}'\,\psi_{\rho y_{i2}})\xi_2 + \psi_{\rho y_{i2}}\eta_{i2},$$

the functions y_i, y_i' occurring explicitly and in the derivatives of ψ_ρ being those belonging to E_{12}. The arc E_{12} is *normal on the sub-interval* $x'x''$ if there exist for it $2n-1$ sets of admissible variations ξ_1^s, ξ_2^s, $\eta_i^s(x)$ such that the matrix

$$\left\| \begin{matrix} \eta_i^s(x') \\ \eta_i^s(x'') \end{matrix} \right\| \tag{2:4}$$

has rank $2n-1$. On account of the relation (2:3) this is the highest rank attainable for a matrix with columns of this sort belonging to an arc that satisfies the equations (2:2) with a set of multipliers $\lambda_\alpha(x)$. For convenience an arc that is normal relative to the end conditions $\psi_\rho=0$ will be designated simply as *normal*.

THEOREM 2:2. *An admissible arc E_{12} that satisfies the necessary condition* I *is normal if and only if there exist for it no set of multipliers $\lambda_\alpha(x)$, not vanishing simultaneously, with which it satisfies equations* (2:2) *and for which the determinant*

$$\begin{vmatrix} 0 & F_{y_i'}(x_1) & 0 & -F_{y_i'}(x_2) \\ \psi_{\rho x_1} + y_{i1}'\psi_{\rho y_{i1}} & \psi_{\rho y_{i1}} & \psi_{\rho x_2} + y_{i2}'\psi_{\rho y_{i2}} & \psi_{\rho y_{i2}} \end{vmatrix} \tag{2:5}$$

vanishes on E_{12}. If E_{12} is normal the constant l_0 defined below can be taken equal to 1, *and its multipliers $\lambda_\alpha(x)$ are then unique.*

To prove the theorem we first notice that the arc E_{12} is normal if and only if there exist for it no set of constants and multipliers l_0, l_ρ, $\lambda_\alpha(x)$ having $l_0=0$ but not vanishing simultaneously with which it satisfies the relations (2:2) and

$$\begin{aligned} l_0(g_{x_1} + y_{i1}'g_{y_{i1}}) + l_\rho(\psi_{\rho x_1} + y_{i1}'\psi_{\rho y_{i1}}) &= 0, \\ l_0 g_{y_{i1}} \qquad\qquad + l_\rho\psi_{\rho y_{i1}} &= F_{y_i'}(x_1), \\ l_0(g_{x_2} + y_{i2}'g_{y_{i2}}) + l_\rho(\psi_{\rho x_2} + y_{i2}'\psi_{\rho y_{i2}}) &= 0, \\ l_0 g_{y_{i2}} \qquad\qquad + l_\rho\psi_{\rho y_{i2}} &= -F_{y_i'}(x_2). \end{aligned} \tag{2:6}$$

This criterion for normality is readily established by the same methods as those used by Bliss for the case when E_{12} is an extremal (V, p. 311). If for a set of multipliers $\lambda_\alpha(x)$ belonging to E_{12} the determinant (2:5) vanishes, then there is a set l_0, l_ρ, $c\lambda_\alpha(x)$ having $l_0=0$ and satisfying the equations (2:6). Hence E_{12} could not be normal. On the other hand if the determinant (2:5) is different from zero for every set of multipliers $\lambda_\alpha(x)$ with which E_{12} satisfies equations (2:2), then there can be no set l_0, l_ρ, $\lambda_\alpha(x)$ with $l_0=0$ satisfying the equations (2:6). Consequently in this case E_{12} is normal. The last statement in the theorem is readily established by the methods used by Bliss for the case when E_{12} is an extremal (V, p. 311).

THEOREM 2:3. *If an admissible arc E_{12} is normal on $x'x''$ and satisfies the equations* (2:2) *with a set of multipliers $\lambda_\alpha(x)$, then these multipliers are unique on the interval $x'x''$ except for a constant factor.*

This is a result of the relation (2:3) which implies that the constants $F_{y_i'}(x')$, $F_{y_i'}(x'')$ are unique except for a constant factor since it is possible to select a matrix (2:4) having rank $2n-1$ on $x'x''$. The multipliers belonging to E_{12} on the interval $x'x''$ are then also unique except for a constant factor since they are completely determined when the set of values $F_{y_i'}(x')$ is specified (IX, p. 680).

3. **The family of extremals.** An *extremal* is an admissible arc with a set of multipliers not vanishing simultaneously

$$y_i = y_i(x), \quad \lambda_\alpha = \lambda_\alpha(x) \qquad (x_1 \leqq x \leqq x_2)$$

which have continuous derivatives $y_i'(x)$, $y_i''(x)$, $\lambda_\alpha'(x)$ and satisfy the Euler-Lagrange equations

$$(3:1) \qquad (d/dx)F_{y_i'} - F_{y_i} = 0, \quad \phi_\alpha = 0.$$

Such an extremal is *non-singular* if the determinant

$$R = \begin{vmatrix} F_{y_i'y_k'} & \phi_{\beta y_i'} \\ \phi_{\alpha y_k'} & 0 \end{vmatrix}$$

is different from zero along it. Along a non-singular extremal E_{12} the equations

$$(3:2) \qquad F_{y_i'}(x, y, y', \lambda) = z_i, \quad \phi_\alpha(x, y, y') = 0$$

can be solved for the variables y_i', λ_α in a neighborhood of the values (x, y, z) on the arc E_{12}. The solution has the form

$$(3:3) \qquad y_i' = P_i(x, y, z), \quad \lambda_\alpha = \Lambda_\alpha(x, y, z),$$

and has continuous partial derivatives of the first two orders since the first members of equations (3:2) have such derivatives. The system of equations (3:1) is now equivalent to the system

$$(3:4) \qquad dy_i/dx = P_i(x, y, z), \quad dz_i/dx = F_{y_i}[x, y, P(x, y, z), \Lambda(x, y, z)].$$

The functions F, P_i, Λ_α satisfy the homogeneity relations

$$(3:5) \qquad \begin{aligned} F(x, y, y', k\lambda) &= kF(x, y, y', \lambda), \\ P_i(x, y, kz) &= P_i(x, y, z), \\ \Lambda_\alpha(x, y, kz) &= k\Lambda_\alpha(x, y, z) \qquad (k \neq 0). \end{aligned}$$

The first of these relations is a consequence of the definition of F. The last two follow from the fact that the two sets

$$[x, y, kz, P(x, y, z), k\Lambda(x, y, z)],$$
$$[x, y, kz, P(x, y, kz), \Lambda(x, y, kz)]$$

satisfy equations (3:2) and must be identical since the solutions P, Λ of these equations are unique when x, y, z are given.

Through every element (x_0, y_0, z_0) in a neighborhood of the set of values (x, y, z) on the extremal E_{12} there passes a unique solution

$$y_i = y_i(x, x_0, y_0, z_0), \quad z_i = z_i(x, x_0, y_0, z_0) \tag{3:6}$$

of equations (3:4) for which the functions y_i, y_{ix}, z_i, z_{ix} have continuous partial derivatives of the first two orders since the second members of equations (3:4) have such derivatives. The functions $y_i(x, x_0, y_0, z_0)$, $kz_i(x, x_0, y_0, z_0)$ are solutions of equations (3:4), on account of the homogeneity properties (3:5), and have the initial values $(x, y, z) = (x_0, y_0, kz_0)$. Since the solutions with these initial values are unique it follows that

$$\begin{aligned} y_i(x, x_0, y_0, kz_0) &= y_i(x, x_0, y_0, z_0), \\ z_i(x, x_0, y_0, kz_0) &= kz_i(x, x_0, y_0, z_0). \end{aligned} \tag{3:7}$$

Since each curve (3:6) has an initial set at $x = x_{10}$ we lose none of them if we replace x_0 by the fixed value x_{10}. Furthermore not all the constants z_{i0} are zero at the initial element of E_{12}. We may therefore renumber the solutions (3:6) so that z_{n0} is different from zero. On account of the homogeneity relations (3:7) it follows that the initial elements (x_{10}, y_0, z_0), (x_{10}, y_0, kz_0) determine the same curves $y_i = y_i(x, x_{10}, y_0, z_0)$. Hence we lose none of these curves if we assign to z_{n0} the fixed value of z_n belonging to E_{12} at the point 1. Let us for convenience rename the constants $y_{10}, y_{20}, \cdots, y_{n0}, z_{10}, \cdots, z_{n-1,0}$ and call them $c_1, c_2, \cdots, c_{2n-1}$ respectively. The family (3:6) then takes the form

$$y_i = y_i(x, c), \quad z_i = z_i(x, c). \tag{3:8}$$

The equations

$$c_i = y_i(x_{10}, c), \quad c_{n+r} = z_r(x_{10}, c), \quad z_{n0} = z_n(x_{10}, c)$$

express the fact that the solutions (3:8) pass through the initial element

$$(x, y_1, \cdots, y_n, z_1, \cdots, z_{n-1}, z_n) = (x_{10}, c_1, \cdots, c_n, c_{n+1}, \cdots, c_{2n-1}, z_{n0})$$

and from them we find by differentiation that the determinant

$$\begin{vmatrix} y_{ic_s} & 0 \\ z_{ic_s} & z_i \end{vmatrix} \tag{3:9}$$

takes the value z_{n0} at $x = x_{10}$. When we substitute the functions (3:8) in

equations (3:3) a set of functions $\lambda_\alpha(x, c)$ is determined, and we have the final result:

THEOREM 3:1. *Every non-singular extremal arc E_{12} is a member of a $(2n-1)$-parameter family of extremals*

$$y_i = y_i(x, c), \quad \lambda_\alpha = \lambda_\alpha(x, c) \qquad (x_1 \leqq x \leqq x_2) \tag{3:10}$$

for special values $(x_1, x_2, c) = (x_{10}, x_{20}, c_0)$. The functions $y_i, y_{ix}, z_i, z_{ix}, \lambda_\alpha$ have continuous first and second partial derivatives in a neighborhood of the values (x, c) defining E_{12}, and for the special values (x_{10}, c_0) the determinant (3:9) *is different from zero.*

4. **The second variation for a normal extremal.** Consider a normal extremal arc E_{12} with ends satisfying the conditions $\psi_\rho = 0$. Let $\xi_1, \xi_2, \eta_i(x)$ be a set of admissible variations along E_{12} satisfying the equations $\Psi_\rho(\xi, \eta) = 0$. It can be shown that there is a one-parameter family of admissible arcs

$$y_i = y_i(x, b), \quad x_1(b) \leqq x \leqq x_2(b), \tag{4:1}$$

satisfying the end conditions $\psi_\rho = 0$, containing E_{12} for $b = 0$, and having $\xi_1, \xi_2, \eta_i(x)$ as its variations along E_{12}(IX, p. 695). The functions $x_1(b)$, $x_2(b)$, $y_i(x, b)$, $y_{ib}(x, b)$ are continuous in a neighborhood of the values (x, b) defining E_{12}, and their derivatives $x_{1b}, x_{1bb}, x_{2b}, x_{2bb}, y_{ix}, y_{ixbb}, y_{ibb}$ have the same property except possibly at the values of x defining the corners of the arc $\eta_i = \eta_i(x)$ $(x_1 \leqq x \leqq x_2)$ in $x\eta$-space.

When the equations

$$\begin{aligned} g(b) &= g[x_1(b), y(x_1(b), b), x_2(b), y(x_2(b), b)], \\ 0 &= \psi_\rho[x_1(b), y(x_1(b), b), x_2(b), y(x_2(b), b)], \\ 0 &= \phi_\alpha[x, y(x, b), y'(x, b)] \end{aligned}$$

are multiplied by constants and multipliers $l_0, l_\rho, \lambda_\alpha(x)$, where l_0, l_ρ are to be determined later and the functions $\lambda_\alpha(x)$ are the multipliers belonging to E_{12}, it is found by suitable additions that

$$\begin{aligned} l_0 g(b) &= G[x_1(b), y(x_1(b), b), x_2(b), y(x_2(b), b)], \\ 0 &= F[x, y(x, b), y'(x, b), \lambda_\alpha(x)], \end{aligned}$$

where $G = l_0 g + l_\rho \psi_\rho$. By differentiating these equations for b it follows further that

$$\begin{aligned} l_0 g'(b) &= (G_{x_1} + y_{i1}' G_{y_{i1}}) x_{1b} + G_{y_{i1}} y_{ib}(x_1) \\ &\quad + (G_{x_2} + y_{i2}' G_{y_{i2}}) x_{2b} + G_{y_{i2}} y_{ib}(x_2), \\ 0 &= F_{y_i} y_{ib} + F_{y_i'} y_{ib}', \end{aligned}$$

and a second differentiation gives for $b=0$

$$l_0 g''(0) = (G_{x_1} + y'_{i1} G_{y_{i1}})x_{1bb} + G_{y_{i1}} y_{ibb}(x_1) \Big|^{b=0}$$

$$(4:2) \qquad + (G_{x_2} + y'_{i2} G_{y_{i2}})x_{2bb} + G_{y_{i2}} y_{ibb}(x_2) \Big|^{b=0}$$

$$+ Q[\xi_1, \eta(x_1), \xi_2, \eta(x_2)],$$

$$(4:3) \qquad 0 = F_{y_i} y_{ibb} + F_{y_i'} y'_{ibb} \Big|^{b=0} + 2\omega(x, \eta, \eta'),$$

where Q is a quadratic form in the variations ξ_1, $\eta_i(x_1)$, ξ_2, $\eta_i(x_2)$ of the family (4:1) along E_{12} and

$$(4:4) \qquad 2\omega(x, \eta, \eta') = F_{y_i y_k}\eta_i\eta_k + 2F_{y_i y_k'}\eta_i\eta_k' + F_{y_i' y_k'}\eta_i'\eta_k'.$$

When equation (4:3) is integrated from x_1 to x_2, it is found with the help of the Euler-Lagrange equations (3:1) that

$$(4:5) \qquad 0 = F_{y_i'} y_{ibb} \Big|_{x_1}^{x_2} + \int_{x_1}^{x_2} 2\omega(x, \eta, \eta')dx.$$

From the hypothesis (c) of § 2, and since E_{12} is normal, we can determine the constants l_0, l_ρ to satisfy equations (2:6) with $l_0 = 1$. Hence by adding equations (4:2) and (4:5) it follows that the second variation I_2 along E_{12} can be expressed in the form

$$(4:6) \qquad I_2 = g''(0) = Q[\xi_1, \eta(x_1), \xi_2, \eta(x_2)] + \int_{x_1}^{x_2} 2\omega(x, \eta, \eta')dx,$$

and this expression must be $\geqq 0$ for every set of admissible variations ξ_1, ξ_2, $\eta_i(x)$ along E_{12} satisfying the conditions $\Psi_\rho(\xi, \eta) = 0$.

Since E_{12} is normal the relation (2:3) and Theorem 2:2 imply that every set of admissible variations ξ_1, ξ_2, $\eta_i(x)$ along E_{12} satisfying the conditions $\Psi_\rho = 0$ also satisfies the equations $\xi_1 = \eta_i(x_1) = \xi_2 = \eta_i(x_2) = 0$. Hence in the expression (4:6) the value of the quadratic form Q is always zero, and we have the following theorem:

THEOREM 4:1. *Along a normal extremal arc E_{12} with ends satisfying the conditions $\psi_\rho = 0$ the second variation is always expressible in the form*

$$I_2 = \int_{x_1}^{x_2} 2\omega(x, \eta, \eta')dx$$

for all admissible variations ξ_1, ξ_2, $\eta_i(x)$ satisfying the equations $\Psi_\rho = 0$, where 2ω is the quadratic form (4:4). If $g(E_{12})$ is to be a minimum for the problem of Mayer as here proposed, then this second variation must be $\geqq 0$ for every set of admissible variations $\eta_i(x)$ satisfying the relations

$$(4:7) \qquad n_i(x_1) = n_i(x_2) = 0.$$

Since the functions $\eta_i(x)$ satisfy the differential equations of variation

$$\Phi_\alpha(x, \eta, \eta') = \phi_{\alpha y_i}\eta_i + \phi_{\alpha y_i'}\eta_i' = 0 \tag{4:8}$$

it is clear that the properties of the second variation suggest a minimum problem which is a problem of Lagrange (cf. VI, p. 16), namely, that of minimizing I_2 in the class of arcs

$$\eta_i = \eta_i(x) \qquad (x_1 \leqq x \leqq x_2) \tag{4:9}$$

satisfying equations (4:8) and passing through the fixed points $(x_1, 0)$, $(x_2, 0)$ in $x\eta$-space as indicated by equations (4:7). One readily verifies that this problem is abnormal since, as was seen in §2, the rank of the matrix (2:4) cannot exceed $2n-1$ on E_{12}. However, by a suitable modification of the end conditions the problem can be made normal. For this purpose we replace the condition that the arc (4:9) passes through the fixed points $(x_1, 0)$, $(x_2, 0)$ in $x\eta$-space by the conditions

$$x_1 - \alpha_1 = \eta_l(x_1) = x_2 - \alpha_2 = \eta_l(x_2) = 0 \qquad (l = 1, \cdots, n;\ l \neq p), \tag{4:10}$$

where p is chosen so that $F_{y_p'}(x_2) \neq 0$. The two sets of end conditions are equivalent since the relation (2:3) implies that $\eta_p(x_2) = 0$ whenever the conditions (4:10) are satisfied.

To prove that the new accessory problem just described is normal we use the fact that since E_{12} is normal there is a determinant of the form $|\Psi_\rho(\xi^\sigma, \eta^\sigma)|$ which is different from zero on E_{12}. The matrix of this determinant is the product of two matrices, the first of which is formed by deleting the first row of the matrix (2:5) and has rank $2n+1$, and the second of which is a matrix having $2n+1$ columns of the form

$$\xi_1^\sigma,\ \eta_i^\sigma(x_1),\ \xi_2^\sigma,\ \eta_i^\sigma(x_2). \tag{4:11}$$

This second matrix must also have rank $2n+1$ if the original determinant is to be different from zero, and the determinant formed from this second matrix by leaving out the row of elements $\eta_p^\sigma(x_2)$ must be different from zero, as one readily sees with the help of the relation (2:3). This last determinant is however one of the form whose non-vanishing insures the normality of the accessory problem with end conditions (4:10).

The Euler-Lagrange equations for the $x\eta$-problem are the equations

$$(d/dx)\,\Omega_{\eta_i'} - \Omega_{\eta_i} = 0, \quad \Phi_\alpha(x, \eta, \eta') = 0, \tag{4:12}$$

where $\Omega(x, \eta, \eta', \mu) = \mu_0\omega + \mu_\alpha\Phi_\alpha$. These equations are known as the *accessory equations* for the original Mayer problem.

THEOREM 4:2. *If the functions $\mu_0=1$, $\mu_\alpha(x)$ are a set of multipliers with which an admissible arc* (4:9) *for the $x\eta$-problem satisfies equations* (4:12), *then every set of functions $\rho_0=1$, $\rho_\alpha(x)$ having this property is of the form $\rho_0=1$, $\rho_\alpha(x)=\mu_\alpha(x)+k\lambda_\alpha(x)$, where the functions $\lambda_\alpha(x)$ are the multipliers for E_{12} and k is an arbitrary constant.*

This follows because if $\rho_0=1$, $\rho_\alpha(x)$ are a second set of multipliers for the arc (4:9), then the differences $\rho_\alpha(x)-\mu_\alpha(x)$ must be multipliers for the original problem and hence be of the form $\rho_\alpha(x)-\mu_\alpha(x)=k\lambda_\alpha(x)$, since E_{12} is normal. This proves the theorem (cf. VI, p. 19).

An admissible arc (4:9) having associated with it a set of multipliers μ_0, $\mu_\alpha(x)$ with which it satisfies equations (4:12) will also satisfy the *transversality condition* for the accessory problem just described if it satisfies the relation $\Omega_{\eta_p'}(x_2)=0$ (IX, p. 693). Since E_{12} is normal and $F_{y_p'}(x_2)\neq 0$ it follows that a solution $\eta_i(x)$, $\rho_0=1$, $\rho_\alpha=\mu_\alpha(x)+k\lambda_\alpha(x)$ of equations (4:12) satisfies the transversality condition $\Omega_{\eta_p'}(x_2)=0$ for a suitably selected value of the constant k.

Let us now assume that E_{12} is also non-singular. Then the determinant R is different from zero along E_{12}, and the equations

$$\Omega_{\eta_i'}(x, \eta, \eta', \mu) = \zeta_i, \quad \Phi_\alpha(x, \eta, \eta') = 0$$

with $\mu_0=1$ can be solved for the variables η_i', μ_α. The solution has the form

$$\eta_i' = H_i(x, \eta, \zeta), \quad \mu_\alpha = M_\alpha(x, \eta, \zeta),$$

and the accessory equations (4:12) with $\mu_0=1$ are now equivalent to the equations

$$\begin{aligned} d\eta_i/dx &= H_i(x, \eta, \zeta), \\ d\zeta_i/dx &= \Omega_{\eta_i}[x, \eta, H(x, \eta, \zeta), M(x, \eta, \zeta)], \end{aligned} \tag{4:13}$$

which are linear and homogeneous in the variables η_i, ζ_i. They have the solution $\eta_i\equiv 0$, $\zeta_i=z_i(x)$, where $z_i(x)$ are the values of the derivatives $F_{y_i'}$ along E_{12}, since the corresponding values $\eta_i\equiv 0$, $\mu_\alpha=\lambda_\alpha$ reduce the first equations (4:12) to the Euler-Lagrange equations (3:1). It is known that for equations (4:13) a set of $2n-1$ solutions u_{is}, v_{is}, whose determinant

$$\begin{vmatrix} u_{is} & 0 \\ v_{is} & z_i \end{vmatrix} \tag{4:14}$$

is different from zero for one value of x, has that determinant different from zero for all values of x. Furthermore every solution (η_i, ζ_i) of equations (4:13) is expressible in the form

$$\eta_i = c_s u_{is}, \quad \zeta_i = c_s v_{is} + k z_i, \tag{4:15}$$

where c_s, k are constants (IV, pp. 153–4). One readily verifies that the columns of the determinant (3:9) are a set of solutions of equations (4:13) like those in the columns of (4:14) (IX, p. 726).

As an immediate consequence of the relation (4:15) it follows that there is one and only one solution (η_i, ζ_i) of equations (4:12) taking prescribed values η_{i0}, ζ_{i0} at a given value x_0. In particular the only solution taking the values $\eta_{i0}=\zeta_{i0}=0$ at $x=x_0$ is the solution $\eta_i\equiv\zeta_i\equiv 0$. Furthermore, since E_{12} is normal the only solution having $\eta_i\equiv 0$ on x_1x_2 is the solution $\eta_i\equiv 0$, $\zeta_i= kz_i(x)$. The same is true on a sub-interval $x'x''$ provided E_{12} is normal on this sub-interval.

5. **The necessary condition of Mayer.** A value $x_3\neq x_1$ is said to define a point 3 *conjugate* to 1 on E_{12} if there exists a solution $\eta_i=u_i(x)$, $\mu_0=1$, $\mu_\alpha=\rho_\alpha(x)$ of equations (4:12) whose functions $u_i(x)$ satisfy the relations $u_i(x_1)=u_i(x_3)=0$ but are not all identically zero on x_1x_3.

IV. THE NECESSARY CONDITION OF MAYER. *Let E_{12} be a non-singular normal extremal arc, normal on every pair of sub-intervals x_1x_3 and x_3x_2. If E_{12} is a minimizing arc for the problem of Mayer as here proposed, then between* 1 *and* 2 *on E_{12} there can be no points* 3 *conjugate to* 1.

If there were a solution $\eta_i=u_i(x)$, $\mu_0=1$, $\mu_\alpha=\rho_\alpha(x)$ of equations (4:12) whose functions $u_i(x)$ vanish at x_1 and x_3 but are not all identically zero on x_1x_3, then for the functions $\eta_i(x)$, μ_0, $\mu_\alpha(x)$ defined by the equations

$$
\begin{aligned}
&\eta_i(x) \equiv u_i(x), \quad \mu_0 = 1, \quad \mu_\alpha(x) \equiv \rho_\alpha(x) \text{ on } x_1x_3,\\
&\eta_i(x) \equiv 0, \qquad\ \ \mu_0 = 1, \quad \mu_\alpha(x) \equiv 0 \qquad \text{on } x_3x_2
\end{aligned}
\tag{5:1}
$$

the second variation I_2 would take the value zero (IX, p. 726). It follows that the arc

$$\eta_i = \eta_i(x) \qquad (x_1 \leqq x \leqq x_2) \tag{5:2}$$

would be a minimizing arc for the $x\eta$-problem since E_{12} is to be a solution of the original problem. Hence there would be associated with the arc (5:2) a function $\Omega=\omega+\mu_\alpha\Phi_\alpha$ with which it would satisfy the accessory equations (4:12), the transversality condition $\Omega_{\eta_p'}(x_2)=0$, and the condition that the derivatives $\Omega_{\eta_i'}(x)$ are continuous on the interval x_1x_2. As was seen above the most general multipliers possible for the functions $\eta_i(x)$ would have the forms $\mu_0=1$, $\mu_\alpha=\rho_\alpha(x)+c\lambda_\alpha(x)$ on the interval x_1x_3 and $\mu_0=1$, $\mu_\alpha=d\lambda_\alpha(x)$ on the interval x_3x_2. On account of the transversality condition $\Omega_{\eta_p'}(x_2)=0$ it is found that $d=0$ since $F_{y_p'}(x_2)\neq 0$. Hence at $x=x_3$ the corner condition would require

$$\Omega_{\eta_i'}(x_3-0) = \omega_{\eta_i'}(x,u,u') + (\rho_\alpha + c\lambda_\alpha)\,\phi_{\alpha y_i'}\,|^{x_3} = 0.$$

It follows that there would exist for the arc (5:2) a set of multipliers $\mu_0=1$, $\mu_\alpha=\rho_\alpha(x)+c\lambda_\alpha(x)$ such that at $x=x_3$ the functions $\zeta_i=\Omega_{\eta_i'}(x,\ u,\ u',\ \rho+c\lambda)$ vanish as well as $\eta_i=u_i$. Hence the functions $\eta_i(x)$, $\zeta_i(x)$ would all vanish identically on x_1x_3 which is not the case, and the theorem is therefore established (cf. VI, p. 18).

6. **The determination of conjugate points.** Consider a non-singular, normal extremal arc E_{12} that is normal on every sub-interval x_1x_3.

THEOREM 6:1. *Let u_{is}, v_{is} be $2n-1$ solutions of equations* (4:13) *whose determinant* (4:14) *is different from zero at $x=x_1$. A value $x_3\neq x_1$ determines a point* 3 *conjugate to* 1 *on E_{12} if and only if the matrix*

$$\left\|\begin{matrix} u_{is}(x_3) \\ u_{is}(x_1)\end{matrix}\right\| \tag{6:1}$$

has rank $<2n-1$.

This theorem is a simple extension of a theorem given by Larew and can be proved by the same methods (VI, p. 20).

If now we select $2n-1$ solutions u_{is}, v_{is} of equations (4:13), as in Theorem 6:1, and such that at $x=x_1$ the functions $u_{is}(x)$ have the values

$$u_{ir}(x_1) = 0, \quad u_{i,n-1+k}(x_1) = \delta_{ik} \qquad (\delta_{ii}=1,\ \delta_{ik}=0 \text{ for } i\neq k),$$

then it is clear that the matrix (6:1) for this set has rank $2n-1$ if and only if the matrix $\|u_{ir}(x_3)\|$ has rank $n-1$. With this in mind we can prove the following theorem:

THEOREM 6:2. *Let u_{ik}, v_{ik} be n solutions of equations* (4:13) *which at $x=x_1$ satisfy the relations*

$$u_{ir}(x_1) = 0, \qquad |v_{ir}(x_1)\ z_i(x_1)| \neq 0,$$
$$u_{in}(x_1) = z_i(x_1), \qquad v_{in}(x_1) = 0.$$

A value $x_3\neq x_1$ determines a point 3 *conjugate to* 1 *on E_{12} if and only if $D(x_3)=0$, where $D(x)=|u_{ik}(x)|$.*

The theorem follows at once from our previous considerations if we show that $D(x_3)$ vanishes if and only if the matrix $\|u_{ir}(x_3)\|$ has rank $<n-1$. If now $D(x_3)=0$, then there exist constants a_k, not all zero, such that $u_{ik}(x_3)a_k=0$. On account of the relation (2:3) for the functions $\eta_i(x)=u_{ik}(x)a_k$ and the values of u_{ik} at $x=x_1$ it follows that

$$0 = z_i(x_3)\; u_{ik}(x_3)a_k = z_i(x_1)\; u_{ik}(x_1)a_k = z_i(x_1)\; z_i(x_1)a_n.$$

Hence $a_n = 0$, and the matrix $\|u_{ir}(x_3)\|$ has rank $< n-1$. The converse is immediate, and the theorem is established.

7. Mayer fields and a fundamental sufficiency theorem. The importance of the introduction of the notion of an $(n+1)$-dimensional field in the space of points $(x, y_1, \cdots, y_n)$ for the problems of Mayer will be seen from the following considerations.

Definition of a Mayer field. A *Mayer field* for the problem considered in this paper is a region $\mathfrak{F}$ in xy-space containing only interior points and having associated with it a set of functions $p_i(x, y), \lambda_\alpha(x, y)$ with the following properties:

(a) they have continuous first partial derivatives in $\mathfrak{F}$;

(b) the sets $[x, y, p(x, y)]$ defined by the points (x, y) in $\mathfrak{F}$ are all admissible;

(c) the integral

$$I^* = \int \{F(x, y, p, \lambda)dx + (dy_i - p_i dx)F_{y_i'}(x, y, p, \lambda)\}$$

formed with these functions is independent of the path in $\mathfrak{F}$.

This definition of a field is precisely the one given by Bliss for the problem of Lagrange except for the form of the function F(IX, p. 730). It should be noted that for the problem of Mayer here discussed the function $F(x, y, p, \lambda)$ vanishes identically in $\mathfrak{F}$, which is not in general true for the problems of Lagrange. Bliss has shown that the solutions $y_i(x)$ of the equations $dy_i/dx = p_i(x, y)$ are extremals with multipliers $\lambda_\alpha(x, y(x))$, called *extremals of the field*. It is clear that the value of I^* is zero along every extremal of the field.

Theorem 7:1. *If E_{12} is a normal extremal arc of a field $\mathfrak{F}$ with ends satisfying the conditions $\psi_\rho = 0$, then there is a neighborhood N of the ends of E_{12} in $(x_1y_1x_2y_2)$-space such that for every admissible arc C_{34} in $\mathfrak{F}$ with ends in N satisfying the conditions $\psi_\rho = 0$ the formula*

$$(7:1) \qquad g(C_{34}) - g(E_{12}) = (1/\lambda_0) \int_{x_3}^{x_4} E[x, y, p(x, y), \lambda(x, y), y']dx$$

holds, where λ_0 is a suitably chosen positive constant,

$$E(x, y, p, \lambda, y') = F(x, y, y', \lambda) - F(x, y, p, \lambda) - (y_i' - p_i)\, F_{y_i'}(x, y, p, \lambda),$$

and the arguments $y_i(x)$, $y_i'(x)$ occurring in the integrand are those belonging to C_{34}.

As a first step in the proof consider the equations

$$g(x_1, y_1, x_2, y_2) = g, \quad \psi_\rho(x_1, y_1, x_2, y_2) = 0.$$

By hypothesis they are satisfied by the set $[x_1, y_1, x_2, y_2, g(E_{12})]$ belonging to E_{12}. Since the determinant (2:1) is different from zero these equations have solutions of the form

$$x_1 = x_1(g), \quad y_{i1} = y_{i1}(g), \quad x_2 = x_2(g), \quad y_{i2} = y_{i2}(g) \tag{7:2}$$

which have continuous second derivatives in a neighborhood of the value $g = g(E_{12})$. Furthermore, in a sufficiently small neighborhood N of the ends of E_{12} the only solutions are those defined by equations (7:2). These equations define two arcs A, B through the ends of E_{12}.

The equations

$$\begin{aligned}
l_0 g_{x_1} + l_\rho \psi_{\rho x_1} &= -\, p_i F_{y_i'}(x, y, p, \lambda)\big|^1, \\
l_0 g_{y_{i1}} + l_\rho \psi_{\rho y_{i1}} &= F_{y_i'}(x, y, p, \lambda)\big|^1, \\
l_0 g_{x_2} + l_\rho \psi_{\rho x_2} &= p_i F_{y_i'}(x, y, p, \lambda)\big|^2, \\
l_0 g_{y_{i2}} + l_\rho \psi_{\rho y_{i2}} &= - F_{y_i'}(x, y, p, \lambda)\big|^2,
\end{aligned}$$

where the variables x_1, y_{i1}, x_2, y_{i2} are replaced by the right members of equations (7:2), determine continuous functions $l_0(g)$, $l_\rho(g)$. When they are multiplied by the differentials dx_1, dy_{i1}, dx_2, dy_{i2} belonging to the arcs A, B and added, it is found that

$$l_0 dg = -\, F_{y_i'}(dy_i - p_i dx)\Big|_1^2. \tag{7:3}$$

In order to compare the values of g for the arcs E_{12} and C_{34} this last equation may be integrated from $g = g(E_{12})$ to $g = g(C_{34})$. By then applying the first law of the mean to the left member, an equation of the form

$$\lambda_0[g(C_{34}) - g(E_{12})] = I^*(A_{13}) - I^*(B_{24}) \tag{7:4}$$

is obtained, where λ_0 is a suitably selected mean value of the function $l_0(g)$ on E_{12}. Since E_{12} is normal we may suppose $l_0 = 1$ on E_{12}, according to the agreement made in §2. Consequently the neighborhood N can be chosen so small that $l_0(g) > 0$ and hence $\lambda_0 > 0$ in N. Furthermore, since I^* is independent of the path in $\mathfrak{F}$ it is clear that

$$I^*(A_{13}) - I^*(B_{24}) = I^*(E_{12}) - I^*(C_{34}) = -\, I^*(C_{34}), \tag{7:5}$$

the last equality being valid since I^* vanishes identically along the extremal E_{12} of the field. The theorem now follows at once from equations (7:4) and (7:5) since, as is easily seen, the value of $-I^*(C_{34})/\lambda_0$ is equal to the value of the second member of equation (7:1).

It is now possible to prove the following important theorem:

THEOREM 7:2. A FUNDAMENTAL SUFFICIENCY THEOREM. *Let a normal extremal arc E_{12} be an extremal of a field $\mathfrak{F}$. Suppose that the ends of E_{12} satisfy the conditions $\psi_\rho=0$ and that there is a neighborhood N of these ends in $(x_1y_1x_2y_2)$-space such that no other extremal of the field has ends in N satisfying the equations $\psi_\rho=0$. If at each point of $\mathfrak{F}$ the condition*

$$E[x, y, p(x, y), \lambda(x, y), y'] > 0$$

holds for every admissible set $(x, y, y')\neq(x, y, p)$, then the neighborhood N can be so restricted that the inequality $g(C_{34})>g(E_{12})$ is true for every admissible arc C_{34} in $\mathfrak{F}$ with ends in N satisfying the conditions $\psi_\rho=0$ and not identical with E_{12}.

To prove this, restrict N so as to be effective as in Theorem 7:1. It follows at once from Theorem 7:1 that the inequality $g(C_{34})\geqq g(E_{12})$ is necessarily satisfied by every admissible arc C_{34} in $\mathfrak{F}$ with ends in N satisfying the conditions $\psi_\rho=0$. The equality sign is appropriate only when the E-function vanishes along C_{34}, that is, only when $y_i'=p_i$ at each point of C_{34}. But in that case C_{34} would be an extremal of the field and would coincide with E_{12} since E_{12} is the only extremal of the field with ends in N satisfying the conditions $\psi_\rho=0$.

8. **An auxiliary theorem.** A normal extremal arc E_{12} is said to satisfy the *Clebsch condition* III' if at each element (x, y, y', λ) on it the inequality

$$F_{y_i'y_k'}\Pi_i\Pi_k > 0$$

holds for every set $(\Pi_1, \cdots, \Pi_n)\neq(0, \cdots, 0)$ which is a solution of the equations $\phi_{\alpha y_i'}\Pi_i=0$. The arc E_{12} satisfies the *Mayer condition* IV' if there is no point 3 conjugate to 1 on E_{12} between 1 and 2 or at 2.

In this section we propose to construct n solutions U_{ik}, V_{ik} of equations (4:13) whose determinant $|U_{ik}(x)|$ is different from zero on x_1x_2 as stated in Theorem 8:1 below. To do this we consider a normal extremal arc E_{12} that is normal on every sub-interval x_1x_3 and satisfies the conditions III', IV' just described. From the condition III' we conclude that E_{12} is non-singular (IX, p. 735).

LEMMA 8:1. *There is an interval $x_1<x\leqq x_1+h$ on which there is no point 3 conjugate to 1 on E_{12}.*

This lemma is readily proved by the methods used by Bliss to establish the corresponding theorem for the problem of Lagrange (IX, pp. 737–740). Bliss makes the stronger assumption that E_{12} is normal on every sub-interval $x'x''$, a restriction which is useful if we wish to show that there are no pairs

of conjugate points whatsoever on E_{12} defined by values $x'x''$ on an interval $x_1 \leqq x \leqq x_1+h$. It can, however, be replaced by the weaker hypothesis that E_{12} is normal on every sub-interval x_1x_3 if we wish to consider only the points 3 conjugate to 1 on E_{12}.

For every pair of solutions (η_i, ζ_i), (u_i, v_i) of equations (4:13) it is known that the expression $\eta_i v_i - u_i\zeta_i$ is a constant. If this constant is zero, then the two solutions are called *conjugate solutions* of equations (4:13). A set of n mutually conjugate solutions of equations (4:13) is said to form a *conjugate system* of solutions.

Consider now the system of solutions u_{ik}, v_{ik} of equations (4:13) defined in Theorem 6:2. One readily verifies that this system forms a conjugate system if the functions $v_{ik}(x)$ are modified so that they satisfy the relation $z_i(x_1) \cdot v_{ik}(x_1)=0$. This can be done by adding to the solution u_{ik}, v_{ik} suitable multiples of the solution $\eta_i \equiv 0$, $\zeta_i = z_i(x)$. Furthermore, since E_{12} satisfies the condition IV' it follows from Theorem 6:2 and Lemma 8:1 that the determinant $|u_{ik}(x)|$ is different from zero on the interval $x_1 < x \leqq x_2$. When the matrices $\|u_{ik}\|$, $\|v_{ik}\|$ are multiplied on the right by the inverse of the matrix $\|u_{ik}(x_2)\|$ a new conjugate system η_{ik}, ζ_{ik} is formed which takes values δ_{ik}, B_{ik} at $x=x_2$, where δ_{ik} equals 0 or 1 according as $i \neq k$ or $i=k$, and $B_{ik}=B_{ki}$. It is clear that the determinant $|\eta_{ik}(x)|$ is also different from zero on the interval $x_1 < x \leqq x_2$. Hence the n-parameter family of solutions of equations (4:13)

$$\eta_i = \eta_{ik}a_k, \quad \zeta_i = \zeta_{ik}a_k \quad (x_1 \leqq x \leqq x_2) \tag{8:1}$$

simply covers a region $\mathfrak{F}$ of points $(x, \eta_1, \cdots, \eta_n)$ whose x-coördinates lie on the interval $x_1 < x \leqq x_2$. Each arc of this family intersects the hyperplane $x=x_2$ in points whose η-coördinates are the parameters a_k defining the arc. Furthermore, on the hyperplane $x=x_2$ the Hilbert integral I_2^* for the $x\eta$-problem defined by the family (8:1) takes the form

$$I_2^* = \int 2\zeta_i d\eta_i = \int 2B_{ik}a_k da_i = \int d(B_{ik}a_i a_k)$$

and hence is independent of the path. It follows that the family (8:1) defines a field $\mathfrak{F}$ (IX, p. 733), and the following lemma is established:

LEMMA 8:2. *If η_{ik}, ζ_{ik} is a conjugate system of solutions taking at $x=x_2$ the values δ_{ik}, B_{ik} just defined, then the determinant $|\eta_{ik}(x)|$ is different from zero on the interval $x_1 < x \leqq x_2$. Furthermore the n-parameter family* (8:1) *of solutions of the accessory equations defines a Mayer field over a region $\mathfrak{F}$ of points $(x, \eta_1, \cdots, \eta_n)$ whose x-coördinates lie on the interval $x_1 < x \leqq x_2$.*

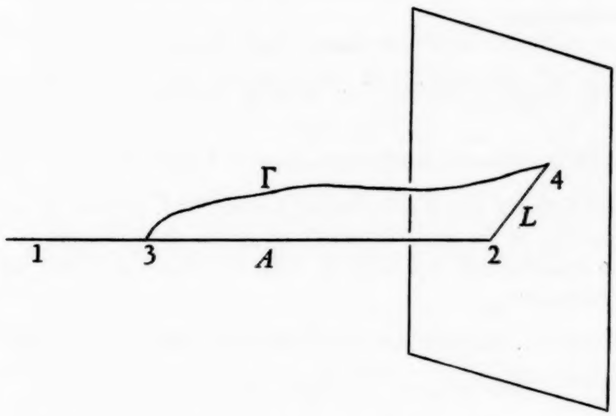
LEMMA 8:3. *For every extremal Γ_{34} for the $x\eta$-problem joining points (x, η)*

$=(x_3, 0)$ *and* $(x, \eta)=(x_2, a)$, *with* $x_1 \leqq x_3 < x_2$, *the relation*

$$I_2(\Gamma_{34}) - B_{ik}a_ia_k \geqq 0 \tag{8:2}$$

holds, where

$$I_2 = \int 2\omega(x, \eta, \eta')dx.$$



Consider first the case when $x_3 > x_1$. According to Lemma 8:2 the Hilbert integral I_2^* for the integral I_2 is independent of the path in $\mathfrak{F}$. Hence

$$\begin{aligned} I_2^*(\Gamma_{34}) &= I_2^*(A_{32}) + I_2^*(L_{24}) \\ &= \int_{L_{24}} 2B_{ik}\eta_i d\eta_k = B_{ik}a_ia_k. \end{aligned} \tag{8:3}$$

Since Γ_{34} is admissible it follows that

$$I_2(\Gamma_{34}) - I_2^*(\Gamma_{34}) = \int_{\Gamma_{34}} E_\Omega dx, \tag{8:4}$$

where E_Ω is the Weierstrass E-function formed for the function 2Ω. By the use of Taylor's expansion one readily verifies that the condition III' on E_{12} implies that $E_\Omega \geqq 0$ along Γ_{34}. Hence from equations (8:3) and (8:4) it is clear that the inequality (8:2) is true whenever $x_3 > x_1$. If now $x_3 = x_1$ then Γ_{34} is an extremal of the field and by direct integration it is found that $I_2(\Gamma_{34}) = B_{ik}a_ia_k$. Hence the lemma is established.

The following theorem gives us the result described at the beginning of this section.

THEOREM 8:1. *Let U_{ik}, V_{ik} be a conjugate system of solutions of equations* (4:13) *having at $x=x_2$ the initial values δ_{ik}, $H_{ik}=B_{ik}-\delta_{ik}$, where δ_{ik}, B_{ik} are the values described above. For such a system the determinant $|U_{ik}(x)|$ is different from zero on the whole interval $x_1\leqq x\leqq x_2$ and $H_{ik}=H_{ki}$.*

In the first place $|U_{ik}(x_2)|=1$. If now $|U_{ik}(x)|$ vanishes for a value x_3 $(x_1\leqq x_3<x_2)$, then there exist constants a_k, not all zero, such that $U_{ik}(x_3)a_k=0$. The equations

$$\eta_i = U_{ik}a_k, \quad \zeta_i = V_{ik}a_k$$

define an arc Γ_{34} as in Lemma 8:3. By direct integration it is found that for this arc

$$I_2(\Gamma_{34}) - B_{ik}a_ia_k = (B_{ik} - \delta_{ik})a_ia_k - B_{ik}a_ia_k = - a_ia_i < 0.$$

This contradicts the result obtained in Lemma 8:3. Hence $|U_{ik}(x_3)|$ is different from zero on the whole interval x_1x_2 as was to be proved.

9. **The construction of a field.** In order to construct a field we need the following theorem:

THEOREM 9:1. *Suppose that an n-parameter family of extremals*

$$y_i = y_i(x, a_1, \cdots, a_n), \quad \lambda_\alpha = \lambda_\alpha(x, a_1, \cdots, a_n) \tag{9:1}$$

is intersected by an n-dimensional manifold

$$x = x_1(a_1, \cdots, a_n), \ y_i = y_i[x_1(a_1, \cdots, a_n), a_1, \cdots, a_n] \tag{9:2}$$

and simply covers a region $\mathfrak{F}$ of xy-space containing only interior points. If the parameter values of the extremal through the point (x, y) are denoted by $a_i(x, y)$, then the region $\mathfrak{F}$ is a field with slope-functions and multipliers

$$p_i(x, y) = y_{ix}[x, a(x, y)], \quad \lambda_\alpha(x, y) = \lambda_\alpha[x, a(x, y)] \tag{9:3}$$

provided that the integral I^* is independent of the path on the n-dimensional manifold (9:2).

This theorem has been established by Bliss for the problem of Lagrange (IX, p. 733). The proof is the same for the problem considered here.

THEOREM 9:2. *If a normal extremal arc E_{12} is normal on every sub-interval x_1x_3 and satisfies the conditions* III', IV', *then E_{12} is a member of an n-parameter family of extremals* (9:1) *whose determinant $|y_{ia_k}|$ is different from zero along E_{12}. Furthermore E_{12} is an extremal arc of a field $\mathfrak{F}$ simply covered by the family.*

To prove this let $W(a_1, \cdots, a_n)$ be a function of the form

$$W(a) = z_{i2}a_i + (1/2)H_{ik}(a_i - y_{i2})(a_k - y_{k2}), \tag{9:4}$$

where the constants y_{i2}, z_{i2} are the values of the functions $y_i(x)$, $z_i(x)$ defining E_{12} at $x=x_2$, and the H_{ik} are the numbers belonging to the conjugate system U_{ik}, V_{ik} defined in Theorem 8:1. When in equations (3:6) the set (x_0, y_{i0}, z_{i0}) is replaced by the set (x_2, a_i, W_{a_i}), an n-parameter family of extremals

$$(9:5)\qquad \begin{aligned} y_i &= y_i(x, x_2, a, W_a) = y_i(x, a), \\ z_i &= z_i(x, x_2, a, W_a) = z_i(x, a) \end{aligned}$$

is defined and contains E_{12} for the special values $a_i = y_{i2}$. The multipliers $\lambda_\alpha(x, a)$ associated with this family are determined by equations (3:3). Furthermore, since each extremal (9:5) defined by parameter values a_i has on it the element (x_2, a_i, W_{a_i}), it follows that $y_{ia_k} = \delta_{ik}$, $z_{ia_k} = W_{a_ia_k} = H_{ik}$ at $x=x_2$. Hence from Theorem 8:1 we conclude that the determinant $|y_{ia_k}|$ is different from zero along each extremal of the family (9:5). This family, therefore, simply covers a neighborhood $\mathfrak{F}$ of E_{12}. Moreover, on the hyperplane $x=x_2$ the Hilbert integral I^* can be expressed in the form

$$I^* = \int F_{y_i'} dy_i = \int W_{a_i} da_i = \int dW$$

and hence is independent of the path. Theorem 9:1 now justifies the theorem that was to be proved.

THEOREM 9:3. *Let a normal extremal arc E_{12} be a member of an n-parameter family of extremals* (9:1) *whose determinant $|y_{ia_k}|$ is different from zero along E_{12}. If the ends of E_{12} satisfy the conditions $\psi_\rho = 0$, then there is a neighborhood N of these ends in $(x_1y_1x_2y_2)$-space such that E_{12} is the only extremal of the family with ends in N satisfying the conditions $\psi_\rho = 0$.*

To prove this let E_{12} be a member of the family (9:1) for the special parameter values (x_{10}, x_{20}, a_0). By hypothesis these values satisfy the equations

$$\psi_\rho(x_1, x_2, a) = \psi_\rho[x_1, y(x_1, a), x_2, y(x_2, a)] = 0.$$

The theorem now follows at once from implicit function theorems if we can show that the matrix

$$(9:6)\qquad \left\| \psi_{\rho x_1} + y_{i1}' \psi_{\rho y_{i1}} \quad \psi_{\rho x_2} + y_{i2}' \psi_{\rho y_{i2}} \quad \psi_{\rho y_{i1}} y_{ia_k}(x_1) + \psi_{\rho y_{i2}} y_{ia_k}(x_2) \right\|$$

has rank $n+2$ on E_{12}. To do this suppose that it had rank less than $n+2$. Then there would exist constants b_1, b_2, c_k, not all zero, such that the relations

$$(\psi_{\rho x_1} + y_{i1}' \psi_{\rho y_{i1}})b_1 + (\psi_{\rho x_2} + y_{i2}' \psi_{\rho y_{i2}})b_2 + \psi_{\rho y_{i1}} y_{ia_k}(x_1)c_k + \psi_{\rho y_{i2}} y_{ia_k}(x_2)c_k = 0,$$
$$F_{y_i'}(x_1) y_{ia_k}(x_1)c_k - F_{y_i'}(x_2) y_{ia_k}(x_2)c_k = 0$$

would hold on E_{12}. The last equation is precisely the relation (2:3) for the

admissible variations $\eta_i = y_{ia_k} c_k$. On account of the normality of E_{12} the determinant (2:5) is different from zero on E_{12}. Hence we would have

$$b_1 = b_2 = y_{ia_k}(x_{10}, a_0)c_k = y_{ia_k}(x_{20}, a_0)c_k = 0.$$

But this is impossible since the determinant $|y_{ia_k}|$ is different from zero along E_{12}. The matrix (9:6) therefore has rank $n+2$ on E_{12}, and the theorem is established.

10. **Sufficient conditions for relative minima.** The condition I is defined in §2, the Clebsch condition III' and the Mayer condition IV' in §8. A normal minimizing arc E_{12} is said to satisfy the Weierstrass condition $\mathrm{II}_{\mathfrak{N}}'$ if at each element (x, y, y', λ) in a neighborhood $\mathfrak{N}$ of those belonging to E_{12} the inequality

$$E(x, y, y', \lambda, Y') > 0$$

holds for every admissible element $(x, y, Y') \neq (x, y, y')$.

THEOREM 10:1. SUFFICIENT CONDITIONS FOR A STRONG RELATIVE MINIMUM. *Let E_{12} be an admissible arc without corners and with ends satisfying the conditions $\psi_\rho = 0$. If E_{12} is normal relative to the end conditions $\psi_\rho = 0$, is normal on every sub-interval x_1x_3 of x_1x_2, and satisfies the conditions* I, $\mathrm{II}_{\mathfrak{N}}'$, III', IV', *then there are neighborhoods $\mathfrak{F}$ of E_{12} in xy-space and N of the ends of E_{12} in $(x_1y_1x_2y_2)$-space such that the inequality $g(C_{34}) > g(E_{12})$ holds for every admissible arc C_{34} in $\mathfrak{F}$ with ends in N satisfying the conditions $\psi_\rho = 0$ and not identical with E_{12}.*

To prove this theorem we first notice that the condition I and the normality of E_{12} imply a unique set of multipliers $\lambda_\alpha(x)$ and constants c_i with which E_{12} satisfies equations (2:2) and for which $l_0 = 1$, as agreed upon in Theorem 2:2. The condition III' implies further that E_{12} is non-singular and hence must be a single extremal arc, since it has no corners (IX, p. 735). According to Theorem 9:2 we now see that E_{12} is an extremal of a field $\mathfrak{F}$ with slope functions and multipliers $p_i(x, y), \lambda_\alpha(x, y)$. It follows that if the field $\mathfrak{F}$ is taken sufficiently small, the values $x, y, p_i(x, y), \lambda_\alpha(x, y)$ belonging to it will lie in so small a neighborhood of the sets (x, y, y', λ) belonging to E_{12} that the condition $\mathrm{II}_{\mathfrak{N}}'$ will imply the inequality

$$E(x, y, p(x, y), \lambda(x, y), y') > 0$$

for every admissible set $(x, y, y') \neq (x, y, p)$ in $\mathfrak{F}$. Theorem 9:3 and the fundamental sufficiency theorem 7:2 now justify the theorem that was to be proved.

Bliss (IX, pp. 736–37) has shown that if an extremal arc E_{12} satisfies the condition III' and is an extremal of a field $\mathfrak{F}$ with slope functions and multipliers $p_i(x, y)$, $\lambda_\alpha(x, y)$, then the inequality

$$E[x, y, p(x, y), \lambda(x, y), y'] > 0$$

holds for every admissible set $(x, y, y') \neq (x, y, p)$ in a neighborhood $\mathfrak{P}$ of the sets (x, y, y') on E_{12}. Hence by arguments like those in the preceding paragraph the following theorem is justified:

THEOREM 10:2. SUFFICIENT CONDITIONS FOR A WEAK RELATIVE MINIMUM. *If an admissible arc E_{12} satisfies all the conditions of the preceding theorem except the condition* II$\mathfrak{N}'$, *then there are neighborhoods $\mathfrak{P}$ of the sets (x, y, y') on E_{12} and N of the end values (x_1, y_1, x_2, y_2) of E_{12} such that the inequality $g(C_{34}) > g(E_{12})$ is true for every admissible arc C_{34} whose elements (x, y, y') are all in $\mathfrak{P}$, whose ends are in N and satisfy the conditions $\psi_\rho = 0$, and which is not identical with E_{12}.*

Suppose now that the functions ψ_ρ are continuous at every pair of distinct or coincident points in a neighborhood of those belonging to E_{12}. Bliss has shown that if the ends of E_{12} are the only pair of distinct or coincident points on E_{12} satisfying the conditions $\psi_\rho = 0$, then for every neighborhood N of the ends of E_{12} in $(x_1y_1x_2y_2)$-space there is a neighborhood $\mathfrak{F}$ of E_{12} in xy-space such that every pair of points (x_1, y_1), (x_2, y_2) in $\mathfrak{F}$ satisfying the conditions $\psi_\rho = 0$ are also in N (XII, p. 267). Hence by suitably restricting the neighborhood $\mathfrak{F}$ of E_{12} in Theorem 10:1 we have the following corollary:

COROLLARY 10:1. *Let E_{12} be an admissible arc satisfying the conditions described in Theorem* 10:1. *If further the ends of E_{12} are the only pair of distinct or coincident points on E_{12} satisfying the conditions $\psi_\rho = 0$, then there is a neighborhood $\mathfrak{F}$ of E_{12} in xy-space such that the inequality $g(C_{34}) > g(E_{12})$ holds for every admissible arc C_{34} in $\mathfrak{F}$ with ends satisfying the conditions $\psi_\rho = 0$ and not identical with E_{12}.*

A similar corollary can be stated for weak relative minima.

BIBLIOGRAPHY

I. Kneser, *Lehrbuch der Variationsrechnung*, Braunschweig, 1900, pp. 227–261.

II. Egorov, *Die hinreichenden Bedingungen des Extremums in der Theorie des Mayerschen Problems*, Mathematische Annalen, vol. 62 (1906), pp. 371–380.

III. Bolza, *Über den anormalen Fall beim Lagrangeschen und Mayerschen Problem mit gemischten Bedingungen und variablen Endpunkten*, Mathematische Annalen, vol. 74 (1913), pp. 430–446.

IV. Goursat, *A Course in Mathematical Analysis*, translated by Hedrick and Dunkel, vol. 2, Part 2.

V. Bliss, *The problem of Mayer with variable end points*, these Transactions, vol. 19 (1918), pp. 305–314.

VI. Larew, *Necessary conditions in the problem of Mayer in the calculus of variations*, these Transactions, vol. 20 (1919), pp. 1–22.

VII. Larew, *The Hilbert integral and Mayer fields for the problem of Mayer in the calculus of variations*, these Transactions, vol. 26 (1924), pp. 61–67.

VIII. Kneser, *Lehrbuch der Variationsrechnung*, 2d edition, Braunschweig, 1925, pp. 240–304.

IX. Bliss, *The problem of Lagrange in the calculus of variations*, American Journal of Mathematics, vol. 52 (1930), pp. 673–742.

X. Morse and Myers, *The problems of Lagrange and Mayer with variable end points*, Proceedings of the American Academy of Arts and Sciences, vol. 66 (1931), pp. 235–253.

XI. Morse, *Sufficient conditions in the problem of Lagrange with variable end conditions*, American Journal of Mathematics, vol. 53 (1931), pp. 517–546.

XII. Bliss, *The problem of Bolza in the calculus of variations*, Annals of Mathematics, vol. 33 (1932), pp. 261–274.

University of Chicago,
Chicago, Ill.

THE TOTAL VARIATION OF $g(x+h)-g(x)$

BY

NORBERT WIENER and R. C. YOUNG*

1. There is a fundamental theorem in the theory of Lebesgue integration that if $f(x)$ be any function integrable Lebesgue, over the interval (a, b), then the integral

$$\int_a^b |f(x+h) - f(x)| \, dx \tag{1}$$

tends to zero with h. This theorem is usually proved by approximating to $f(x)$ in terms of continuous functions, for which the property is obvious.

The integral (1) is the total variation in (a, b) of the function $F(x+h)-F(x)$, where

$$F(x) = \int^x f(t)dt$$

and the quoted property is equivalent to the following statement:

(I) *If $F(x)$ be any absolutely continuous function in (a, b), the total variation of*

$$F(x+h) - F(x)$$

tends to zero with h.

This result no longer holds if we substitute for $F(x)$ any non-absolutely continuous function $G(x)$ of bounded variation. Indeed, it provides a necessary and sufficient condition for absolute continuity.† This fact may be first rendered plausible by taking the simplest case of a discontinuous function of bounded variation, and bearing in mind that a general continuous function of bounded variation is always a limit of simple discontinuous ones. If we assume for instance

$$g(x) = \alpha \text{ for } x \leqq c,$$
$$g(x) = \beta \text{ for } c < x,$$

* Presented to the Society, October 29, 1932; received by the editors April 21, 1932.

† This result was proved, by an entirely different method, by A. Plessner, Journal für Mathematik, vol. 160 (1929), pp. 26–32. A study of the total variation of $g(x+h)-g(x)$ when $g(x)$ is constant in the complementary intervals of Cantor's set is contained in an article by Hille and Tamarkin, American Mathematical Monthly, vol. 36 (1929), pp. 255–264. The existence of these two papers was called to our attention after the present paper had been completed. We notice also an erroneous statement in a paper by J. M. Whittaker, Proceedings of the Edinburgh Mathematical Society, (2), vol. 1 (1929), p. 232 (Lemma 1).

we have, for $h>0$,

$$g(x+h)-g(x)=0 \qquad \text{for } x \leqq c-h \text{ and } x>c,$$
$$g(x+h)-g(x)=\beta-\alpha \text{ for } c-h<x\leqq c$$

and over any interval (a, b) containing c and $c+h$ internally, the total variation of $g(x+h)-g(x)$ is the sum of the absolute values of its jumps, viz.

$$2\left|\beta-\alpha\right| = 2\left|\int_a^b dg(x)\right| = \left|\int_a^b dg(x)\right| + \left|\int_a^b dg(x+h)\right|.$$

In this special case we may then write, for every h, and every interval (a, b), $a\neq c$, $b\neq c$,

$$\int_a^b \left| d[g(x+h)-g(x)]\right| = \int_a^b \left| dg(x)\right| + \int_a^b \left| dg(x+h)\right|. \tag{2}$$

Actually this is the relation which we shall show (§3) to hold for any singular function $g(x)$ of bounded variation, if not for all h, at any rate for almost all. By a singular function we mean a function of bounded variation whose derivative is zero almost everywhere, of which the above discontinuous $g(x)$ provides a very special example.

An arbitrary function of bounded variation is the sum of an absolutely continuous function and a singular function of bounded variation, uniquely defined and called the singular part of the original function. Thus the proof of the above statement will carry with it, as an immediate consequence of I, the following result:

(II) *If $g(x)$ be any function of bounded variation with singular part $\gamma(x)$ continuous at the end points a, b,* we have*

$$\begin{aligned}\overline{\lim_{h\to 0}} \int_a^b \left| d[g(x+h)-g(x)]\right| &= \overline{\lim_{h\to 0}} \int_a^b \left| d[\gamma(x+h)-\gamma(x)]\right| \\ &= 2\int_a^b \left| d\gamma(x)\right|.\end{aligned} \tag{3}$$

This statement is of course only a rough corollary of the result for singular functions and absolutely continuous functions. More precise consequences to be born in mind are

(i) the left-hand equality in (3) holds not merely for h tending to 0 continuously, but for h tending to 0 through any discontinuous sequence;

* As an unessential restriction reducing $\int_{a+0}^{b+0}|d\gamma|$ to $\int_a^b|d\gamma|$.

(ii) the right-hand equality in (3) holds similarly for discontinuous approach of h to 0, if a certain set of measure 0 be avoided by h. To obtain any intermediary or the lower value of the considered limits, we have to take subsequences of this set of measure 0.

Examples will be constructed (§§4–6) to show what various possibilities exist with regard to the exceptional set of values of h. It will be seen that this set may be more than countable, or countable, or entirely absent. In particular therefore the limits in (3) may be unique limits. On the other hand, the examples will also show that the lower limits corresponding to the upper limits in (3) may in other cases have any lesser non-negative value. The case in which the lower limit is 0 would seem to have particular interest, and it might be useful to examine in greater detail the particular sequences of h_n tending to 0, for which this limit is obtained. It seems a question for instance whether a non-absolutely continuous function $g(x)$ exists for which $h_n=1/n$ would provide a sequence of this kind, or more generally, any other special sequence for which h_n/h_{n+1} is bounded.

2. We use the following simple lemmas.

LEMMA 1. *Given two functions $g_1(x)$, $g_2(x)$ of bounded variation, of which $g_1(x)$ has derivative zero, $g_1'(x)=0$, on the set complementary to a given set H_1 in (a, b). Then we have*

$$(4)\qquad \int_a^b |d[g_1(x)-g_2(x)]| = \int_a^b |dg_1| + \int_a^b |dg_2| - 2\theta\int_{H_1}|dg_2|,$$

where $0\leqq\theta\leqq 1$.*

We have

$$\int_{CH_1}|dg_2| - \int_{CH_1}|dg_1| \leqq \int_{CH_1}|d[g_2(x)-g_1(x)]|$$
$$\leqq \int_{CH_1}|dg_2| + \int_{CH_1}|dg_1|.$$

Since

$$\int_{CH_1}|dg_1| = 0,$$

we have

* This lemma has been stated initially in a less general form. The authors are indebted to Dr. S. Saks for the suggestion of extending their proof to the present, essentially more general, situation. We refer to de la Vallée Poussin, *Intégrales de Lebesgue, Fonctions d'Ensemble, Classes de Baire*, Paris, 1916, pp. 90–95, as to the notion of the total variation over a set, and as to some formulas which we are using here.

$$\int_{CH_1} | d[g_2 - g_1] | = \int_{CH_1} | dg_2 | .$$

On the other hand, over the set H_1 itself,

$$- \int_{H_1} | dg_2 | \leqq \int_{H_1} | d[g_2 - g_1] | - \int_{H_1} | dg_1 | \leqq \int_{H_1} | dg_2 | .$$

Hence, by addition,

$$\int_a^b | dg_2 | - 2 \int_{H_1} | dg_2 | \leqq \int_a^b | d[g_2 - g_1] | - \int_a^b | dg_1 | \leqq \int_a | dg_2 |$$

giving (4) as required.

LEMMA 2. *Let E_h be a variable set depending on a parameter h, and such that each point x belongs to E_h at most for a set of measure* 0 *of values of h*. Then for each function $g(x)$ of bounded variation,*

$$\int_{E_h} | dg | = 0$$

for almost all values of h. (More precisely, the exceptional values of h form a set of measure 0 *depending on g.)*

This is immediate from the theory of change of order of integration in a repeated Stieltjes integral.† Let $G(x)$ be the indefinite total variation of $g(x)$:

$$G(x) = \int^x | dg |$$

and $E(x, h)$ the characteristic function of E_h, equal to 1 in E_h, and 0 elsewhere, for each value of h. We have then

$$V(h) = \int_{E_h} | dg | = \int_{-\infty}^{\infty} E(x, h) dG(x).$$

Integrating this with respect to h, we get

$$\int_{-\infty}^{\infty} V(h) dh = \int_{-\infty}^{\infty} dh \int_{-\infty}^{\infty} E(x, h) dG(x) = \int_{-\infty}^{\infty} dG(x) \int_{-\infty}^{\infty} E(x, h) dh,$$

* In other words, E_h is the section y=constant h, of a plane set E whose sections x=constant all have measure 0. Such a set E has plane measure 0, and it is well known that E_h must then have measure 0 for almost all h. The present Lemma states that E_h has also measure 0 with respect to any function of bounded variation, for almost all h, and is an immediate adaptation of the classical result.

† See, e.g., L. C. Young, *The Theory of Integration*, Cambridge Tracts, No. 21, p. 41 (Theorem IV).

where, by the hypothesis,

$$\int_{-\infty}^{\infty} E(x, h)dh = 0 \text{ for each } x,$$

as the measure of the set of values of h for which x belongs to E_h. Thus

$$\int_{-\infty}^{\infty} V(h)dh = 0,$$

and hence the non-negative function $V(h)$ vanishes except at most in a set of measure 0.

In our application of this Lemma, the set E_h will be the h-translation of a fixed set E_0 of measure 0, that is, the set of points x such that $x-h$ belongs to E_0. It has the characteristic function

$$E_h(x) = E_0(x + h).$$

For fixed x this still represents a set of measure 0 in h, and so the hypothesis of the Lemma is fulfilled. Thus we have

COROLLARY. *Given any function of bounded variation $g(x)$, and any set E_0 of measure* 0, *the total variation of g over the h-translation of E_0 vanishes except at most for a set of values of h of measure* 0.

3. We have now immediately our

THEOREM. *If $g(x)$ be a function of bounded variation constant in the complementary intervals of a closed set H of measure* 0, *we have, for almost all h,*

$$\int_a^b | d[g(x + h) - g(x)]| = \int_a^b | dg(x + h)| + \int_a^b | dg(x)| . \tag{5}$$

For, $g(x+h)$ is then constant in the complementary intervals of the h-translation H_h of H, so by Lemma 1, for

$$g_2(x) = g(x + h), \; g_1(x) = g(x),$$

the two sides of (5) differ by at most

$$2\int_{H_h} | dg |$$

and by Lemma 2, in the Corollary form, this vanishes for almost all h.

In constructing examples, it is simplest to consider periodic functions with interval of periodicity (a, b), and assume for instance $a=0$, $b=1$. Then (5) takes the form

$$\int_0^1 | d[g(x + h) - g(x)]| = 2\int_0^1 | dg(x)|. \tag{5'}$$

In that case also it suffices to consider only positive values of h, since $g(x+h)-g(x)$ is then still periodic with the same period as g, and so

$$\int_0^1 |d[g(x+h)-g(x)]| = \int_0^1 |d[g(x)-g(x-h)]|\,.$$

Furthermore we need only consider monotonic functions, with $\int_0^1 dg=1$.

4. The classical example of a singular function of bounded variation is that of the monotone function constant in the complementary intervals of Cantor's typical ternary set* and representing an even mass distribution over this set, in the interval $(0, 1)$. If x be expressed as a ternary fraction

$$(6) \qquad x = \cdot\alpha_1\alpha_2\cdots\alpha_i\cdots = \sum_{i=1}^{\infty}\alpha_i 3^{-i},\ \alpha_i = 0,\ 1 \text{ or } 2,$$

and α_n be the first (if any) of its digits equal to 1, the function is defined at x to have the value

$$(7) \qquad g(x) = \sum_{i=1}^{n-1}\alpha_i 2^{-i-1} + \alpha_n 2^{-n},$$

or if there be no digit 1 in (6), then

$$(7') \qquad g(x) = \sum_{i=1}^{\infty}\alpha_i\, 2^{-i-1}.$$

This function is a particular example of a class of functions possessing the following property; this is most easily described by introducing the expression "ternary interval of order n" to designate specifically the open intervals of length $1/3^n$ whose left-hand end points are the terminating ternary fractions of at most n digits, all even:

$$\cdot\alpha_1\alpha_2\cdots\alpha_n\,0\,0\,0\cdots,\ \alpha_i = 0 \text{ or } 2.$$

Each such interval is contained in exactly one of lower order, and contains exactly 2^{r-n} intervals of higher order $r>n$. The intervals of given order are of course mutually exclusive and non-abutting.

Property (A). *For each ternary interval (α, β) of order n, and each ternary interval (α', β') of order $(n+l)$ contained in it,*

$$\left|\frac{g(\beta')-g(\alpha')}{g(\beta)-g(\alpha)}\right| \leqq \delta$$

where l is a positive integer and δ a constant less than 1, *independent of the particular interval.*

* The set of all non-terminating ternary fractions with even digits only, together with the terminating fractions whose last digit at most is odd.

Such functions share with the ordinary measuring function $m(x)=x$ a property relative to fractions with prescribed digits which we shall use in the form of

LEMMA 3. *If $g(x)$ be any function, constant in the complementary intervals of Cantor's ternary set H_0, and possessing the property* (A), *and if E be a subset of H_0 in which* (*in the ternary scale*)

$$x = \cdot\alpha_1\alpha_2 \cdots \alpha_i \cdots \tag{6}$$

has fixed prescribed digits (0 *or* 2) *for an infinity of indices,*

$$\alpha_{n_1}, \alpha_{n_2}, \cdots, \alpha_{n_i}, \cdots, n_1 < n_2 < \cdots \to \infty$$

(*the other digits being arbitrary,* 0 *or* 2), *then*

$$\int_E |dg| = 0.$$

Consider the ternary intervals of order n_k and left-hand end points

$\cdot\beta_1\beta_2 \cdots \beta_{n_1-1}\alpha_{n_1}\beta_{n_1+1} \cdots \beta_{n_2-1}\alpha_{n_2}\beta_{n_2+1} \cdots \beta_{n_k-1}\alpha_{n_k}$, α_{n_i} as above, $\beta_i = 0$ or 2,

and let σ_k denote their sum-set. Then $E \subset \sigma_k$ for each k (actually $E = \lim_{k\to\infty}\sigma_k$). Also

$$\int_{\sigma_k} |dg| \leqq \delta \int_{\sigma_{k-l}} |dg| . \tag{8}$$

This is because each interval of σ_k is in a different ternary interval of order $n_k - l$ contained in σ_{k-l}, and by property (A) contributes not more than δ times the variation over this interval to the total variation over σ_k. Thus if $[k/l]$ be the integral part of k/l,

$$\int_E |dg| \leqq \int_{\sigma_k} |dg| \leqq \delta^{[k/l]} \int_0^1 |dg| \to 0 \text{ with } 1/k$$

since $\delta < 1$.

This lemma is applied in conjunction with another relating purely to the ternary set H_0:

LEMMA 4. *To each non-terminating ternary fraction $h = \cdot a_1 a_2 \cdots a_i \cdots$ there corresponds an infinite sequence of indices*

$$n_1 < n_2 < \cdots < n_i < \cdots \to \infty$$

and two specifications for the whole sequence of digits $\{c_{n_i}\}$ (0 *or* 2), *which a ternary fraction $x = \cdot c_1c_2 \cdots c_i \cdots$ must certainly satisfy if it is to belong to both H_0 and its h-translation H_h.*

This means that the common part H_0H_h is the sum of two sets of the kind considered in Lemma 3. What we prove precisely is as follows:

(i) If h has an infinite number of odd digits

$$a_{n_i} = 1,\ n_1 < n_2 < \cdots < n_i < \cdots \to \infty$$

then the corresponding digits of x must be alternately 0 and 2, i.e.

$$c_{n_1} = c_{n_3} = c_{n_5} = \cdots = 0 \text{ or } 2,$$
$$c_{n_2} = c_{n_4} = c_{n_6} = \cdots = 2 \text{ or } 0.$$

(ii) If h has only a finite number of odd digits,

$$a_i \neq 1 \text{ for } i > N,$$

and (r_i) are the indices for which $a_i=0$, (s_i) those for which $a_i=2$, after the Nth, then in x we must have either

$$c_{r_1} = c_{r_2} = \cdots = c_{r_i} = \cdots = 0$$

or

$$c_{s_1} = c_{s_2} = \cdots = c_{s_i} = \cdots = 2.$$

We deduce this from the following obvious facts, true for any h, terminating or not:

Given in the ternary scale

$$h = \cdot a_1a_2 \cdots a_i \cdots \qquad (a_i = 0, 1 \text{ or } 2),$$
$$x = \cdot b_1b_2 \cdots b_i \cdots \qquad (b_i = 0 \text{ or } 2),$$
$$x + h = \cdot c_1c_2 \cdots c_i \cdots \pmod 1 \qquad (c_i = 0 \text{ or } 2),$$

we have

(a) $$\text{if } a_{n_1} = a_{n_2} = 1,\ a_i \neq 1 \text{ for } n_1 < i < n_2,$$

then either

$$b_{n_1} = 0,\quad b_{n_2} = 2 \ (\text{and } c_{n_1} = 2,\ c_{n_2} = 0),$$

or the same with n_1 and n_2 interchanged;

(b) $$\text{if}\quad a_r = b_r = 0\ (= c_r),\ a_s = b_s = 2\ (= c_s),$$

then $a_i=1$ for some index between r and s (implying $|r-s|>1$).

In (a) we consider two consecutive odd digits in h, and affirm that the two corresponding digits in x (and $x+h$) cannot be both 0 or both 2. In fact $b_{n_2}=2$ would imply that we had in the formal addition $x+h$ to carry 1 from the n_2th place right back to the n_1th, where it would compound to 2 with a_{n_1}, and imply $b_{n_1}=0$ if c_{n_1} is to be even. And similarly $b_{n_2}=0$ implies $b_{n_1}=2$.

In (b) we consider two places in each of which h and x have the same

digits, but different in the two places, and we conclude that between those two places h must have at least one digit $=1$. We may assume that between those two places no further coincidences occur, and we see at once that if the formal addition $x+h$ is to yield only even digits between those two places, and h had none $=1$ there, these digits in $x+h$ would all be 0 with 1 to carry or all 2, with nothing to carry, and a 1 would appear in the sum in the earlier of the two places r and s.

From Lemmas 3 and 4, we deduce that, *for any function* $g(x)$ *of period* 1, *constant in the complementary intervals of Cantor's set* H_0 *and possessing the above property* (A), *we must have, whenever* h *is a non-terminating ternary fraction,*

$$\int_{H_h} |dg| = \int_{H_0 \cdot H_h} |dg| = 0.$$

From this and Lemma 1, we deduce (as in the proof of our principal theorem), that *for all such functions* $g(x)$,

$$\int_0^1 | d[g(x+h) - g(x)] | = 2 \int_0^1 | dg |$$

whenever h *is a non-terminating ternary fraction.*

The exceptional values of h thus belong to the countable set of terminating ternary fractions.

5. In the case of Cantor's function (7, 7′), property (A) holds with $l=1$ and $\delta=\frac{1}{2}$, and equality sign. This provides therefore an instance of a function $g(x)$ for which the exceptional values of h are at most countable. It may be seen moreover that in this case every terminating ternary fraction is an exceptional value of h. For this it suffices to remark that (5) can certainly not hold when $g(x+h)-g(x)$ is constant in an interval in which neither $g(x)$ nor $g(x+h)$ is constant. And if, when

$$h = \cdot a_1 a_2 \cdots a_i \cdots a_n,$$

we choose (as we can)

$$x_0 = \cdot b_1 b_2 \cdots b_i \cdots b_n, \quad \cdot b_i = 0 \text{ or } 2,$$

so that x_0+h has also only digits 0 or 2, then the ternary interval of order n and left-hand end point x_0 satisfies this condition.* If we go into the question more nearly, in order to investigate the lower limit of

$$\int_0^1 | d[g(x+h) - g(x)] | \quad \text{as} \quad h \to 0,$$

* Cf. also Hille and Tamarkin, loc. cit., p. 261.

we find that for our present function, whereas

$$\int_0^1 |d[g(x+h)-g(x)]| = 2$$

for h non-terminating, we have for h terminating of exactly n digits,

$$1 \leqq \int_0^1 |d[g(x+h)-g(x)]| \leqq 2(1-2^{-n}), \tag{9}$$

with actual equality on either side for suitable values of h and each n. For let

$$h = \cdot a_1a_2 \cdots a_n, \quad a_i = 0, 1 \text{ or } 2, \quad a_n \neq 0;$$

and suppose to fix the ideas that $a_n = 2$. Then every ternary interval of order n and left-hand end point

$$\cdot b_1b_2 \cdots b_n, \quad b_i = 0 \text{ or } 2,$$

with $b_n = 2$, translates through h into an interval in which g is constant, that is to say is itself an interval in which $g(x+h)$ is constant. Similarly if $a_n = 1$ and we take $b_n = 0$. So over half the ternary intervals of order n, the relation (5) certainly holds. Similarly it holds for all the intervals of length 2^{-n} and left-hand end points $\cdot b_1b_2 \cdots b_k$ with $b_k = 1, k \leqq n$, over which $g(x)$ is constant. There only remain the other half of the ternary intervals of order n, over which $g(x)$ has variation 1/2, and $g(x+h)$ at most 1/2. So in this case the right-hand side of (5′) cannot exceed the left by more than 1.

The other inferences follow also very simply. For instance

$$a_n = a_{n-k_1} = \cdots = a_{n-k_r} = 1, \quad a_i = 0 \text{ or } 2 \text{ for } n - k_j < i < n - k_{j-1},$$
$$j = 1, 2, \cdots, r+1, \quad k_0 = 0, \quad k_{r+1} = n,$$

according as j is odd or even, give some of the (everywhere dense set of) values of h for which equality occurs on the right of (9), while

$$a_i = 2 \text{ for } i < n, \ a_n = 1 \ \text{ or } \ a_i = 0 \text{ for } i < n, \ a_n = 2$$

give the only values for which equality occurs on the left.

6. The classical monotone function just discussed expresses the total mass in $(0, x)$ due to mass unity, evenly distributed among the ternary intervals of order n, for each n, and its constancy in each complementary interval of the ternary set H_0 expresses the fact that there is no mass in these complementary intervals.

We now modify this construction so that the function is constant not only in the complementary intervals of H_0, but also in a set (everywhere dense on H_0) of the ternary intervals themselves, and show that we can choose these additional intervals of constancy (in a more than countable number of ways)

so that the resulting functions satisfy (5) for every terminating ternary fraction h. The functions will all possess the property (A), and therefore satisfy (5) for non-terminating values of h, and so we shall have our example of functions for which h has no exceptional values, (5) being always true.

The definition is obtained inductively by supposing the mass distribution effected in the ternary intervals of order $2n$ and now dividing the mass of each such interval equally among *three* (instead of all) of the four ternary intervals of order $2(n+1)$ which it contains. The choice of these three intervals may be effected in four different ways, and as there will be exactly 3^n ternary intervals of order $2n$ actually containing mass, this means that we have $4\cdot 3^n$ different modes of distribution to choose from at each stage. In any case, each ternary interval of order $n+2$ will contain at most 1/3 of the mass of the interval of order n containing it (viz. either 0 or exactly 1/3 of it).

As the mass of any interval is the increment of the monotone function $g(x)$ representing the ultimate mass in $(0, x)$, we see that each function so obtained possesses the property (A) of page 332 with $l=2$, $\delta=1/3$.

The construction provides moreover a more than countable set of functions of this kind, for all of which (5) therefore holds whenever h is a non-terminating ternary fraction. By eliminating a certain minority of these functions, we may arrange so that (5) holds also for each of the remaining countable set of values of h. The simplest way of doing this is to associate with each of these values of h, i.e., terminating of say k digits, an infinite sequence of indices $>k$:

$$N_i = N_i^{(h)} \to \infty$$

such that

$$N_i^{(h)} \neq N_j^{(h')} (\text{all } i, j)$$

whenever $h' \neq h$ (this can always be done in any number of ways*), and to stipulate that for each $n = N_i^{(h)}$, two ternary intervals of order $2n$ which are h-translations one of the other (for that h) should never both be intervals devoid of mass. For instance if

$$h = \cdot a_1 a_2 \cdots a_k, \qquad a_i = 0, 1 \text{ or } 2, \; a_k \neq 0,$$

and

* For instance if $\{h_i\}$ be the considered set of values of h in any countable order, and $\{p_i\}$ the sequence of primes in increasing order, we can associate with each h_i the sequence of integers $p_i p_{i+n}$, $n = n_0, n_0+1, \cdots$.

$$x = \cdot b_1 b_2 \cdots b_k b_{k+1} \cdots b_{2n}, \quad b_i = 0 \text{ or } 2,$$

is the left-hand end point of a ternary interval of order $2n$ devoid of mass, and if $x+h$ is still a left-hand end point of a ternary interval of order $2n$, i.e.

$$x + h = \cdot c_1 c_2 \cdots c_k b_{k+1} \cdots b_{2n}, \; c_i = 0 \text{ or } 2,$$

then this latter interval must not be devoid of mass. Since $2n > k+1$, the two ternary intervals considered are necessarily in two different intervals of order $2(n-1)$, so that at the stage $n-1$ in our construction we can certainly make our choice (in at least 3^{3^n} different ways) so that the condition be fulfilled for that n. By the fact that to each index n that we have particularly to take into consideration corresponds only one h (with reference to which the restriction is made), a question of incompatibility for the different values of h cannot arise, and we are certainly left with a majority of the functions considered. For this residue, we see that for each h, (5) holds outside at most two of the ternary intervals of order $2n$ in each ternary interval of order $2(n-1)$ for each $n = N_i^{(h)}$. Thus if σ_i represent the intervals of this order $2n$ for which (5) may possibly not hold (more precisely in which neither $g(x)$ nor $g(x+h)$ is constant), we have as in the proof of Lemma 2 (inequality (8)),

$$\int_{\sigma_i} | dg | \leqq \delta \int_{\sigma_{i-1}} | dg |, \; \delta = \tfrac{2}{3},$$

and hence this variation tends to 0 with $1/i$. Thus (5) holds in the complementary intervals of a set over which $g(x)$ has total variation 0, that is, holds absolutely, for the arbitrary considered terminating h.

7. The object of rarifying the exceptional values of h as much as possible was achieved by an increased concentration of the unit mass distributed over our interval (0, 1), introducing additional intervals of constancy for $g(x)$. Conversely, we can multiply the exceptional set of values of h by breaking up the intervals of constancy, and diffusing the total mass more over the whole interval.

By this means we shall obtain a function with a more than countable set of exceptional h, and at the same time we shall find our example of a case in which the lower limit corresponding to (3) is 0.*

Let $n_1, n_2, \cdots$ be a sequence of integers increasing so rapidly that $\sum_1^\infty 1/n_k$ converges. Let x be any number in (0, 1). It is a theorem due to Cantor that x may always be expressed in the form

$$x = \frac{m_1}{n_1} + \frac{m_2}{n_1 n_2} + \frac{m_3}{n_1 n_2 n_3} + \cdots, \tag{10}$$

* The actual construction is adapted from a suggestion of Mr. A. S. Besicovitch.

where $m_1<n_1$, $m_2<n_2$, $\cdots$, and that this representation is unique, save for the ambiguity

$$\frac{1}{n_1n_2\cdots n_k}=\frac{n_{k+1}-1}{n_1\cdots n_{k+1}}+\frac{n_{k+2}-1}{n_1\cdots n_{k+2}}+\cdots,$$

which gives a non-terminating alternative to any terminating expression of a number. Let us assume that every n_k is even, equaling $2\nu_k$. If x is expressed as in (10), let

$$f(x)=\frac{m_1/2}{\nu_1}+\frac{m_2/2}{\nu_1\nu_2}+\frac{m_3/2}{\nu_1\nu_2\nu_3}+\cdots,$$

if every m_k is even, and

$$f(x)=\frac{m_1/2}{\nu_1}+\frac{m_2/2}{\nu_1\nu_2}+\cdots+\frac{m_{n-1}/2}{\nu_1\nu_2\cdots\nu_{n-1}}+\frac{[m_n/2]+1}{\nu_1\nu_2\cdots\nu_n},$$

if m_n is the first odd m_k. Clearly $f(x)$ is a monotone function, constant over every interval

$$\left(\frac{2\mu_k+1}{n_1n_2\cdots n_k},\ \frac{2\mu_k+2}{n_1n_2\cdots n_k}\right)\text{ where }\mu_k<\nu_k.$$

These intervals for any given value of k fill up half the line, and for $k=1, 2, 3, \cdots, K$ fill up a set of intervals of measure $1-2^{-K}$.

Let h be defined as

$$\frac{a_1}{n_1}+\frac{a_2}{n_1n_2}+\cdots+\frac{a_k}{n_1n_2\cdots n_k}+\cdots,$$

where every a_k is 0 or 2. We have

$$f(x+h)-f(x)=\frac{a_1}{2\nu_1}+\frac{a_2}{2\nu_1\nu_2}+\frac{a_3}{2\nu_1\nu_2\nu_3}+\cdots$$

whenever

$$2\leqq m_r\leqq n_r-2,\ 2\leqq m_{r+1}\leqq n_{r+1}-2,\cdots,$$

where r is the first index for which $a_k\neq 0$. The total variation of $f(x)$ over the set of points for which $m_k<2$ does not however exceed $1/\nu_k$, and the same is true of the total variation of $f(x+h)$ over those intervals. A similar result applies to the range $n_k-m_k<2$. Thus the total variation of $f(x+h)-f(x)$ over these intervals does not exceed

$$(11)\qquad 4(1/\nu_r+1/\nu_{r+1}+\cdots+1/\nu_{r+k}+\cdots)$$

which is hence an upper bound for the total variation of $f(x+h)-f(x)$ over $(0, 1)$. The total variation of $f(x)$ over the same interval is 1. Thus we have here a set of values of h of the power of the continuum including arbitrarily small values, for which

$$\int_0^1 | d[f(x+h) - f(x)] | \neq 2 \int_0^1 | df(x) | ,$$

and the total variation on the left-hand side tends to zero (since (11) does so) when h tends to zero through this set.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY,
CAMBRIDGE, MASS.
GIRTON COLLEGE,
CAMBRIDGE, ENGLAND

ON A SERIES OF INVOLUTORIAL CREMONA TRANSFORMATIONS OF SPACE DEFINED BY A PENCIL OF RULED SURFACES*

BY

VIRGIL SNYDER

1. **Introduction.** The transformations here discussed are of interest on account of the peculiar manner in which the fundamental elements enter; these singularities have not been mentioned in the existing literature.

Consider a pencil of surfaces $|F_n|$ having a rational curve r to multiplicity $n-2$ as part of its base curve. Let the points M of r and the surfaces of the pencil $|F|$ be in $(1, k)$ correspondence. A point P fixes a surface $F(p)$ of the pencil and this in turn a point M on r. The line PM meets $F(p)$, in one residual point P'. The relation between P, P' is an involutorial birational transformation of space. The case in which the surfaces F are of order n, and the base curve an $n-2$ fold line and a single residual curve, have been considered by Carroll.† The case in which the curve r is of order >1 and the residual curve simple, including the (ruled) quartics through a double cubic curve r, has been discussed by Black.‡

The present paper discusses the possible cases in which every surface of the pencil is ruled. With the exception of the pencil of quartics through a double cubic curve and of the cubics with a common double directrix, the surfaces of the pencil are of order $2n+m$; the base curve consists of a line p to multiplicity $2(n-1)+m$ and a double curve r of order n meeting the line in $n-1$ points. Apart from the multiple directrix line and the double curve, the residual base of the pencil consists entirely of generators; each surface of the pencil accounts for m parasitic lines.

The new transformations include a number of well known types, but also many new ones in which m, n, k may each take any positive integral value.

2. **Equations.** The rational curve r defined parametrically by

$$x_1 = f\lambda, \quad x_2 = f\mu, \quad x_3 = g, \quad x = h,$$

wherein f, g, h are binary forms in λ, μ of degree $n-1, n, n$ respectively, meets the line $p: x_1 = 0, x_2 = 0$ in $n-1$ points. Let a point $M \equiv (0, 0, z_3, z_4)$ on p, and a point (λ, μ) on r be in $(2, m)$ correspondence defined by $z_3^2 u + z_3 z_4 v + z_4^2 w = 0$,

* Presented to the Society, December 29, 1932; received by the editors, October 10, 1932.

† American Journal of Mathematics, vol. 54 (1932), pp. 707–717.

‡ These Transactions, vol. 34 (1932), pp. 795–810.

wherein u, v, w are binary forms of degree m in λ, μ. Lines joining corresponding points generate a ruled surface. Since $x_1\mu = x_2\lambda$, if we write

$$f(x_1, x_2)x_3 - g(x_1, x_2) = fx_3 - g \equiv \xi, \text{ and } fx_4 - h \equiv \eta,$$

its equation has the form

$$F \equiv \xi^2 u + \xi\eta v + \eta^2 w = 0. \tag{1}$$

The surface is of order $2n+m$; it contains r to multiplicity 2 and p to multiplicity $2(n-1)+m$. From every point of p issue m generators apart from p, and the line p counts for $2(n-1)$ generators, two for each point of intersection of r and p.

Let $F'=0$ be of the same kind, in which u, v, w are replaced by u', v', w' of the same degree m, and f, g, h be the same as before. Let surfaces of the pencil

$$lF' - l'F = 0 \tag{2}$$

and points M of p be in $(k, 1)$ correspondence defined by $z_3\phi_4(l, l') - z_4\phi_3(l, l') = 0$, each ϕ_i being a fixed binary form of degree k.

A point $(y) \equiv (y_1, y_2, y_3, y_4)$ in space uniquely fixes the surface of the pencil (2) passing through it: $l = F(y)$, $l' = F'(y)$, $z_3 = \phi_3(F(y), F'(y))$, $z_4 = \phi_4(F(y), F'(y))$. A point (x) on the line joining (y) to M has coordinates of the form

$$\rho x_1 = \tau y_1, \quad \rho x_2 = \tau y_2, \quad \rho x_3 = \sigma\phi_3 + \tau y_3, \quad \rho x_4 = \sigma\phi_4 + \tau y_4.$$

For the point (y') in which the line meets $F_y = 0$ again,

$$\tau = (\xi z_4 - \eta z_3)f[\xi\{z_3(u'v) + z_4(u'w)\} + \eta\{z_3(u'w) + z_4(wv')\}],$$
$$\sigma = -(\xi z_4 - \eta z_3)[\xi^2(u'v) + 2\xi\eta(u'w) + \eta^2(wv')],$$

wherein $(u'v) \equiv u'v - uv'$, etc.

The relation between the points (y) and (y') is an involutorial Cremona transformation I of space. The factor $\xi z_4 - \eta z_3$ divides out of the transformation. When z_3, z_4 are fixed, $\xi z_4 - \eta z_3 = 0$ represents a cone containing r and having M on p for vertex. It meets each surface of the pencil (2) belonging to M in the generators passing through M, and in the base lines p and r of the pencil.

We may now write, after removing the factor,

$$\tau = f[z_3\{\xi(u'v) + \eta(u'w)\} + z_4\{\xi(u'w) + \eta(v'w)\}],$$
$$\sigma = -[\xi\{\xi(u'v) + \eta(u'w)\} + \eta\{\xi(u'w) + \eta(v'w)\}].$$

Under I every surface of the pencil is transformed into itself. The invariant points are the points of contact of tangents from M to the k surfaces belonging to M. But since all the surfaces of the pencil (1) are ruled, if any point is fixed,

the entire generator passing through the point is fixed. These lines generate the surface $\sigma=0$. The surface $\tau=0$ is the image of the line p. It consists of two parts, $f=0$, images of the $n-1$ points of r on p, and of a ruled surface, image of the other points of p. The generators of each surface of (1) passing through its associated point M are fundamental lines of the second kind. When (y) is chosen on any such line, the point (y') is not defined, but is the whole line passing through the point (y). As M describes r these lines describe a ruled surface R defined by

$$R:\ \xi\phi_4 - \eta\phi_3 = 0, \tag{3}$$

which plays the most important part in the transformation.

3. **Transformation of the pencil of planes (p).** Every plane through p is transformed into itself. In each such plane the involution is of order $2k+2$; it is of the non-perspective Jonquières type, having the isolated point on r as fundamental point of order $2k+1$. The plane meets each surface of (1) in p and in two lines which are interchanged by I. Among them are two lines which are invariant point by point; the locus of these lines is the surface $\sigma=0$. The class is k, as is also evident from the definition of the space transformation. The image of the point L on r consists of $2k+1$ lines, all belonging to $\tau=0$. The images of the lines of the plane are curves of order $2k+2$ having $2k+1$ common tangents at the fundamental point L. There are also $2k+1$ simple fundamental points of the plane transformation on p. The lines joining these to L on r are all generators of the surface R. Of the $2(k+1)(n-1)+(k+2)m$ tangent planes through p which belong to a general surface of the web of conjugates of the planes of space, $2(k+1)(n-1)+km$ are fixed for every surface of the web; of these $2(k+1)(n-1)$ are the planes of $f=0$ each counted to multiplicity $2(k+1)$, and the km other ones are tangent planes of R, defined by $y_3\phi_4-y_4\phi_3=0$. The $2m$ variable tangent planes of the conjugate of the plane $(ax)=0$ are defined by

$$x_3\{a_3(u'v)+a_4(u'w)\}+x_4\{a_3(u'w)+a_4(v'w)\}=0.$$

Moreover, all the surfaces of the web also touch each other along every one of the $2k+1$ sheets through r. These are defined by $y_3z_4-y_4z_3=0$. There is simple contact along τ and r, including all the sheets of each, and f counted $2(k+1)$ times.

When a generator in any plane π through p passes through M, it is parasitic, that is, a fundamental line of the second kind. The point M is then the conjugate of the other generator in π. Thus, to each point M of p correspond mk lines, fundamental of the first kind, and the line p is itself multiply parasitic for every point on it.

4. **Residual base elements.** The base of the pencil (1) consists of r to multiplicity 2, of p to multiplicity $2(n-1)+m$, and of $4(n-1)+4m$ generators. Of these, $4(n-1)$ consist of tangency along p in the planes of $f=0$ and $4m$ are generators not coincident with p. Let g_i be one of these latter base generators, and π_i the plane p, g_i. Corresponding to every point M on p are k residual generators in π_i associated with M, all belonging to a pencil of lines with vertex on r. As M describes p, these lines generate π_i in such manner that the complete image of g_i is π_i counted k times.

There are $k+1$ positions of M for which the residual generator on some surface of the pencil (1) passes through it; these are all parasitic lines and are base lines of the web of surfaces conjugate to the planes of space. The composite surface $\tau=0$ consists of the planes $f=0$, images of the $n-1$ points of r on p, and of a ruled surface of order $(2k+1)n+(k+2)m$, having p to multiplicity $(2k+1)n+(k+2)m-(k+1)$ and r to multiplicity $2k+1$. The locus $\sigma=0$ of invariant points is a ruled surface of order $2(n+m)$, has p to multiplicity $2(n-1)+2m$, and r to multiplicity 2. Since all the fundamental elements are included in $\tau=0$, $\sigma=0$, the complete configuration can now be accounted for.

5. **Table of characteristics.** The images of planes and of fundamental elements can now be expressed by the following table:

$$s_1 \sim s_{2(k+1)n+(k+2)m} : p^{2(k+1)(n-1)+(k+2)m} r^{2k+1} 4mg_i 4kmg';$$

$$p \sim \tau_{2(k+1)n+(k+2)m-1} : p^{2(k+1)(n-1)+(k+2)m} r^{2k+1} 4mg_i 4kmg';$$

$$r \sim R_{(2k+1)n+km} : p^{(2k+1)(n-1)+km} r^{2k+1};$$

$$g_i \sim \pi_i p g_i k g';$$

$$\sigma \equiv \sigma_{2(n+m)} : p^{2(n-1)+2m} r^2 4mg_i 4kmg';$$

$$J \equiv \tau^2 R^2 [(vu')(wv') - (wv')^2] f^2.$$

All the surfaces of the web have the same tangent planes along all the $2k+1$ sheets through r. These are the tangent planes of the ruled surface R of fundamental lines of the second kind. The line p illustrates Montesano's theorem for exceptional fundamental lines of the second kind.*

Thus, the complete intersection of any two surfaces of the web consists of the curve conjugate of a line, of the basis elements p, r each to the multiplicity indicated, simple contact along each sheet of $\tau=0$ along p, of R along r, of $4m(k+1)$ lines and of each sheet of f counted $2(k+1)$ times.

* D. Montesano, *Sulla teoria generale delle corrispondenze birazionali dello spazio*, Rendiconti della Accademia dei Lincei, (5), vol. 27_1 (1918), pp. 396–400 and pp. 438–441; and (5), vol. 30_1 (1921), pp. 447–451.

6. **Types not in the preceding category.** The preceding list includes all possible types for arbitrary values of n and m, but for particular values others may appear.

A pencil of quadrics and an arbitrary line p, not a basis line, lead at once to a series of transformations which include one discussed by Montesano.*

The line p may be replaced by any rational curve. The congruence of bisecants of the base C_4 is left invariant.

The next case is that of a pencil of cubic ruled surfaces having a common double directrix. The residual is then a rational quintic r_5 meeting the double directrix in four points. The point M is now on r_5. The residual section of a plane through d and M consists of a generator through M on each of the k surfaces associated with M. Each of these lines is parasitic. Every point P of d is invariant except for the plane containing a generator through P.

Let d be $x_3=0$, $x_4=0$. Then $F_3 \equiv x_3^2 u + x_3 x_4 v + x_4^2 w = 0$, $u=0$ etc. being planes, $F_3' \equiv x_3^2 u' + \cdots = 0$.

Let $\mu x_4 - \lambda x_3 = 0$. Then the parametric equations of r_5 are

$$\rho x_1 = r(\lambda, \mu), \quad x_2 = s(\lambda, \mu), \quad x_3 = \lambda f(\lambda, \mu), \quad x_4 = \mu f(\lambda, \mu),$$

in which r and s are quintics and f a quartic binary form. Since f is a factor of τ, the image of d includes the $4k$ surfaces of the pencil associated with the points in which r_5 meets d. The image of the line d is the surface $R \equiv x_3 \phi_4(F, F') - x_4 \phi_3(F, F') = 0$, of order $3k+1$ containing d to multiplicity $2k+1$, and τ to multiplicity k. Every generator is parasitic, hence R also appears as a factor in the transformation. Given a point P on r_5. The image of P on $F(M)$ is the residual point P' in which the line PM meets $F(M)$. As M describes r_5, this line describes a rational quartic cone, and the locus of P' is a curve of order $4k+N$ having P to multiplicity N. The tangent plane to $F(M)$ at P meets r_5 twice at P and in three other points K. Conversely, given K, then the line KP and the tangent t to r_5 at P uniquely fix a tangent plane to F, hence the point M. The $(1, 3k)$ correspondence between M and K on a rational carrier has $3k+1$ points of coincidence, hence P is of multiplicity $3k+1$, and the curve is of order $7k+1$. In addition, any point M on r_5 is transformed into the conics through it in the tangent plane to each $F(M)$ at M, residual to the generator g_M. This makes the complete curve of order $9k+1$. Hence r appears to multiplicity $9k+1$ on the surfaces conjugate to the planes of space in the involution. This can also be seen directly from the equations of the transformation: Consider a general point P on d. It remains invariant on

---

* *Su una classe di trasformazioni razionali ed involutorie dello spazio di genere arbitrario n e di grado* $2n+1$, Giornale di Matematiche di Battaglini, vol. 31 (1893), pp. 36–50.

every $F(M)$ of the pencil except when PM is the generator g_M on $F(M)$. This happens only for the four positions of M on d. The surface R meets any F of the pencil in d counted $4k+2$ times, in r counted k times, and in a single generator. Let D be one of the points (d, r). Any plane through d meets $F_{(M)}$ in d^2 and a generator g. The image of the generator is D. Hence $F_{(M)}$ is the image not of the line d but of the point D on d.

$$
\begin{aligned}
s_1 &\sim s_{27k+4}: d^{18k+1} r^{9k+1} 4D^{18k+1};\\
d &\sim R_{3k+1}: d^{2k+1} r^k 4D^{2k+1};\\
r &\sim \tau_{27k+3}: d^{18k+1} r^{9k+1} 4D^{18k+1};\\
D &\sim kF_3: d^{2k} r^k 4D^{2k}.
\end{aligned}
$$

The contacts along d and r are as in the general case.

Various particular cases may arise, when r_5 is composite, consisting of one or more generators and a residual curve. When the generator is taken as a projector curve, the result is included in the preceding category; when the residual curve is the locus of M, the image of each generator is a surface of order 6, found as in the general case. The order of the transformation is lowered by unity for each base generator, since the plane d, g will divide out.

A third particular case is that in which the pencil of quartics have a common double cubic curve. This has been fully treated by Black.*

The space cubic may be replaced by a line and a conic meeting it in one point, or by three lines, one of which meets each of the other two, which are skew. The common transversal is a double generator. When either double directrix is used as projector, this is included in the general case, whether the basal double generator exists or not. When the double generator is used as projector, it is included among those treated by Carroll, the residual basis line now consisting of four simple generators, all skew to the double one. But this case offers differences that warrant a more detailed treatment, since no other generators meet g, hence there is no surface of parasitic lines apart from the planes d, g and d', g.

Let $d \equiv x_1=0,\ x_2=0,\ d' \equiv x_3=0,\ x_4=0$ be the two double directrices, and $g \equiv x_1=0,\ x_3=0$ the double generator. The equation of the quartic has the form $F_4 \equiv x_1^2 u + x_1 x_2 x_3 v + a x_2^2 x_3^2 = 0$, wherein u is a binary quadratic form in x_2, x_4, v is linear in x_2, x_4, and a is constant.

Let u', v', a' define F_4' having the same double elements. The pencil $lF' - l'F = 0$ is then associated with the point $M \equiv (0, 0, z_3, z_4)$ of g by the relation $z_3\phi_4(l, l') - z_4\phi_3(l, l') = 0$, wherein as before ϕ_i is a binary form in l,

* Loc. cit.

l' of degree k. The residual base of the pencil consists of four generators g_i, which meet d and d' but do not meet g. The pencil contains two composite surfaces, one consisting of the plane $x_1=0$ and a ruled cubic containing d' as double directrix, and the other of $x_2=0$ and a ruled cubic containing d as double directrix. These planes both divide out, each reducing the order of the transformation by one.

The conjugate of the double generator g is generated by the residual conics to surfaces of $|F|$ at M. It is a surface of order $8k+4$, has d, d' and g each to multiplicity $4k+2$, and the simple basic generators g_i each to multiplicity $2k+1$. This surface is ruled.

The surface $\sigma=0$ of invariant points is of order $4k+5$, has d, d', g each to multiplicity $2k+1$, and the lines g_i each to multiplicity $k+1$.

The image of g_i in the plane Mg_i consists of k nodal cubic curves. As M describes g, these curves generate a surface of order $4k+1$, containing d, d', g each to multiplicity $2k$, g_i to multiplicity $k+1$, and the other basic generators to multiplicity k. The images of d, d' are the planes (d, g), (d', g) respectively. The transformation is now completely defined.

CORNELL UNIVERSITY,
ITHACA, N. Y.

NOTES ON THE THEORY AND APPLICATION OF FOURIER TRANSFORMS. I-II*

BY

R. E. A. C. PALEY AND N. WIENER

Introduction

We propose to publish under the above title a series of notes. The results are of a varied nature, but the methods we employ are very similar and consist, roughly speaking, in conformally mapping the unit circle into a half plane, and considering the Fourier transforms of functions defined on the boundary of the half plane. The notes may be read independently.

I. On a theorem of Carleman

1. The chief object of this note is to give a simple proof of the following theorem which is substantially the same as one due to Carleman.† Let $A_0=1, A_1, \cdots, A_\nu, \cdots$ be a set of positive numbers, and let C_A denote the set of functions defined in the interval $(-\infty, \infty)$, infinitely many times differentiable in that range, and satisfying the inequalities

$$(*) \qquad \int_{-\infty}^{\infty} |f^{(\nu)}(x)|^2 dx \leqq B^\nu A_\nu^2 \qquad (\nu = 0, 1, 2, \cdots),$$

where B is a constant which may depend on $f(x)$. We say that the class C_A is quasi-analytic if a function of C_A is defined completely over $(-\infty, \infty)$ by the values of its derivatives $f^{(\nu)}(x)$ $(\nu=0, 1, 2, \cdots)$ at a single point x_0, or, what is the same thing, if the equations

$$f^{(\nu)}(x_0) = 0 \qquad (\nu = 0, 1, 2, \cdots),$$

together with the condition $f(x) \subset C_A$, imply that $f(x)$ vanishes identically. The theorem is the following:

Theorem I. *A necessary and sufficient condition that C_A should be quasi-analytic is that the integral*

$$(1) \qquad \int_0^\infty \log\left(\sum_{\nu=0}^{\infty} \frac{x^{2\nu}}{A_\nu^2}\right) \frac{dx}{1+x^2}$$

* Presented to the Society, February 25, 1933; received by the editors January 19, 1933. [Mathematical science has suffered an irreparable loss in the untimely death of R. E. A. C. Paley. He was killed on April 7, 1933, at the age of twenty-five, in an accident that occurred during a skiing excursion near Banff, Alberta. J. D. Tamarkin.]

† T. Carleman, *Les Fonctions Quasi-Analytiques*, 1926. We have slightly modified Carleman's definition of C_A. We consider $\int_{-\infty}^{\infty} |f^{(\nu)}(x)|^2 dx$ instead of $\max_{-1\leqq x\leqq 1} |f^{(\nu)}(x)|$, the problem in this form being more adaptable to our attack, but the difference is not at all essential.

should diverge, or, what is the same thing, that the least non-increasing majorant of the series

$$\sum_{\nu=0}^{\infty} 1/(A_\nu)^{1/\nu}$$

should diverge.

The equivalence of the two conditions has been established by Carleman in his book.* In this paper we shall concern ourselves only with the first one.

2. We begin by proving the following theorem.

THEOREM II. *Let $\phi(x)$ be a real non-negative function not equivalent to zero, defined for $-\infty<x<\infty$, and of integrable square in this range. A necessary and sufficient condition that there should exist a real- or complex-valued function $F(x)$ defined in the same range, vanishing for $x \geqq x_0$ for some number x_0, and such that the Fourier transform $G(x)$ of $F(x)$ should satisfy $|G(x)|=\phi(x)$, is that*

$$\int_{-\infty}^{\infty} \frac{|\log \phi(x)|}{1+x^2}\, dx < \infty. \tag{2}$$

We observe that the theorem is similar to one due to de la Vallée Poussin.† He concerns himself with the Fourier coefficients of a periodic function all of whose derivatives vanish at some fixed point. Here we demand rather more of the function, and we undertake actually to fix the modulus of the transform, subject of course to the convergence of (2).

3. Suppose first that the integral (2) converges. We write for $z=x+iy$, $y>0$,

$$\lambda(z) = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\log \phi(x')y}{(x-x')^2+y^2}\, dx',$$

which is harmonic in the half plane $y>0$. Let $\mu(z)$ be its conjugate, and write

$$h(z) = \exp\,[\lambda(z) + i\mu(z)].$$

It is well known, by an argument of the Fatou type, that, for almost all x,

$$\lim_{y\to 0} \lambda(x+iy) = \log \phi(x),$$

or, what is the same thing,

$$\lim_{y\to 0} |h(x+iy)| = \phi(x).$$

We observe first that, by the convexity property,

* Carleman, loc. cit., pp. 50ff.

† Carleman, loc. cit., pp. 76 and 91.

$$| h(x+iy) | = e^{\lambda(z)} \leqq \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\phi(x')y}{(x-x')^2+y^2}dx'$$

$$\leqq \frac{1}{\pi}\left\{\int_{-\infty}^{\infty}\phi(x')^2dx'\right\}^{1/2}\left\{\int_{-\infty}^{\infty}\left(\frac{y}{(x-x')^2+y^2}\right)^2dx'\right\}^{1/2}$$

and tends to zero as $y\to\infty$, uniformly in x. This shows that $h(x+iy)$ is uniformly bounded in any half-plane $y \geqq y_0 > 0$.

Next

$$\int_{-\infty}^{\infty} | h(x+iy) |^2 dx \leqq \frac{1}{\pi}\int_{-\infty}^{\infty}dx \int_{-\infty}^{\infty}\frac{\phi(x')^2 y}{(x-x')^2+y^2}dx'$$

$$= \frac{1}{\pi}\int_{-\infty}^{\infty}\phi(x')^2dx'\int_{-\infty}^{\infty}\frac{ydx}{(x-x')^2+y^2}$$

$$= \int_{-\infty}^{\infty}\phi(x')^2dx',$$

and is therefore uniformly bounded in y. Now let $0<y_0<y<y_1$. Cauchy's theorem gives

$$-2\pi i h(x+iy)$$

$$= \int_{-N}^{N}\frac{h(x'+iy_0)}{(x-x')+i(y-y_0)}dx' - \int_{-N}^{N}\frac{h(x'+iy_1)}{(x-x')+i(y-y_1)}dx'$$

$$+ \int_{y_0}^{y_1}\frac{h(N+iy')}{(x-N)+i(y-y')}dy' - \int_{y_0}^{y_1}\frac{h(-N+iy')}{(x+N)+i(y-y')}dy'.$$

Making first N and then y_1 tend to infinity in the last formula we obtain

$$h(x+iy) = -(2\pi i)^{-1}\int_{-\infty}^{\infty}\frac{h(x'+iy_0)}{(x-x')+i(y-y_0)}dx'. \tag{3}$$

Now let $H_y(\xi)$ denote the Fourier transform

$$\underset{A\to\infty}{\text{l.i.m.}}\ (2\pi)^{-1/2}\int_{-A}^{A}h(x+iy)e^{ix\xi}dx$$

of $h(x+iy)$. Since the Fourier transform of

$$-(2\pi i)^{-1}[x+i(y-y_0)]^{-1}$$

is

$$(2\pi)^{-1/2}e^{(y-y_0)\xi}$$

for negative ξ, and vanishes for positive ξ, it follows that

$$H_y(\xi) = \begin{cases} (2\pi)^{1/2} H_{y_0}(\xi) \cdot (2\pi)^{-1/2} e^{(y-y_0)\xi}, & \xi < 0, \\ 0, & \xi > 0. \end{cases}$$

Thus we have $H_y(\xi) = 0$ for $\xi > 0$ and all positive y, and this gives

$$H_{y_0}(\xi) = \begin{cases} e^{(y_0-y)\xi} H_y(\xi), & \xi < 0, \\ 0, & \xi > 0. \end{cases}$$

Let us keep y fixed and make y_0 tend to zero. Since

$$\int_{-\infty}^{\infty} | h(x + iy_0) |^2 dx$$

is bounded, it follows that

$$\int_{-\infty}^{0} | H_y(\xi) e^{(y_0-y)\xi} |^2 d\xi \tag{4}$$

is bounded and increasing as y_0 decreases to zero, and thus the integral (4) tends to a limit as $y_0 \to 0$, and $H_y(\xi) e^{(y_0-y)\xi}$ tends to $H_y(\xi) e^{-y\xi}$ in the mean of order 2. The Fourier transform of the function which coincides with $H_y(\xi) e^{(y_0-y)\xi}$ for $-\infty < \xi < 0$, and which vanishes for $\xi > 0$, is $h(x+iy_0)$, and hence $h(x+iy_0)$ tends in mean of order 2 to a function $G(x)$ as $y_0 \to 0$. We have shown that the Fourier transform of $h(x+iy_0)$ (with y_0 fixed and positive) vanishes for $\xi > 0$, and it follows that the same is true of the Fourier transform $F(\xi)$ of $G(x)$. We have already seen that $|G(x)| = \phi(x)$.

Now suppose that $F(x')$ vanishes for $x' > x_0$, where we may suppose without loss of generality that $x_0 = 0$. We are to show that the integral (2) converges. We write

$$G(x) = \underset{N\to\infty}{\text{l.i.m.}}\ (2\pi)^{-1/2} \int_{-N}^{N} F(x') e^{-ixx'} dx',$$

$$\psi(z) = \underset{N\to\infty}{\text{l.i.m.}}\ (2\pi)^{-1/2} \int_{-N}^{N} F(x') e^{-izx'} dx', \quad \Im z > 0.$$

The function $\psi(z)$ is readily seen to be analytic in the half plane $\Im z > 0$. Suppose that we invert the half plane $\Im z > 0$ into the circle $|\zeta| < 1$, $\zeta = re^{i\theta}$, and that $G(x)$ becomes $\Gamma(e^{i\theta})$ and $\psi(z)$ becomes $\gamma(\zeta)$. Then it is easily seen that

$$\int_{-\pi}^{\pi} | \Gamma(e^{i\theta}) |^2 d\theta = 2 \int_{-\infty}^{\infty} \frac{| G(x) |^2}{1 + x^2} dx,$$

so that Γ certainly is of class L^2. Also a simple computation shows that if $re^{i\phi}$ is the inverse of $x' + iy'$, then

$$(2\pi)^{-1}\int_{-\pi}^{\pi}\Gamma(e^{i\theta})\frac{1-r^2}{1-2r\cos(\theta-\phi)+r^2}d\theta = \pi^{-1}\int_{-\infty}^{\infty}G(x)\frac{y'dx}{(x-x')^2+y'^2}$$
$$= \pi^{-1}\int_{-\infty}^{\infty}\frac{y'dx}{(x-x')^2+y'^2}\ \underset{N\to\infty}{\text{l.i.m.}}\ (2\pi)^{-1/2}\int_{-N}^{0}F(\xi)e^{-ix\xi}d\xi$$
$$= \lim_{N\to\infty}\pi^{-1}\int_{-\infty}^{\infty}\frac{y'dx}{(x-x')^2+y'^2}(2\pi)^{-1/2}\int_{-N}^{0}F(\xi)e^{-ix\xi}d\xi$$
$$= \lim_{N\to\infty}\pi^{-1}\int_{-N}^{0}\frac{F(\xi)d\xi}{(2\pi)^{1/2}}\int_{-\infty}^{\infty}\frac{e^{-ix\xi}y'dx}{(x-x')^2+y'^2}$$
$$= \lim_{N\to\infty}(2\pi)^{-1/2}\int_{-N}^{0}F(\xi)e^{-ix'\xi+y'\xi}d\xi = \psi(x'+iy') = \gamma(re^{i\phi}),$$

so that γ is in fact the Poisson integral of $\Gamma(e^{i\theta})$. Then

$$(5)\qquad \begin{aligned}&(2\pi)^{-1}\int_{-\pi}^{\pi}\log^+|\gamma(re^{i\theta})|\,d\theta\\ &\leqq (2\pi)^{-1}\int_{-\pi}^{\pi}|\gamma(re^{i\theta})|^2d\theta \leqq (2\pi)^{-1}\int_{-\pi}^{\pi}|\Gamma(e^{i\theta})|^2d\theta.\end{aligned}$$

It is known by a theorem of Ostrowski* and Nevanlinna that the boundedness of the integral (5) implies that of the integral

$$(2\pi)^{-1}\int_{-\pi}^{\pi}\big|\log|\gamma(re^{i\theta})|\big|d\theta.$$

Since finally $\log|\gamma(re^{i\theta})|$ tends almost everywhere to $\log|\Gamma(e^{i\theta})|$ we have

$$(2\pi)^{-1}\int_{-\pi}^{\pi}\big|\log|\Gamma(e^{i\theta})|\big|\,d\theta < \infty,$$

and inverting again to the half plane this implies that

$$\int_{-\infty}^{\infty}\frac{\big|\log|G(x)|\big|}{1+x^2}dx < \infty,$$

which is the same as (2).

4. Returning now to the proof of Carleman's theorem we observe first that if the integral (1) converges, and

$$\phi(x)^2 = (10)^{-1}(1+x^2)^{-1}\left[\sum_{\nu=0}^{\infty}\frac{x^{2\nu}}{A_\nu^2}\right]^{-1},$$

* See, e. g., A. Ostrowski, *Über die Bedeutung der Jensenschen Formel für einige Fragen der komplexen Funktionentheorie*, Acta Szeged, vol. 1 (1923), pp. 80–87.

then

$$\int_{-\infty}^{\infty} \frac{|\log \phi(x)|}{1+x^2}\,dx < \infty, \quad \int_{-\infty}^{\infty} \phi(x)^2 dx < \infty,$$

so that we can find a function $F(x) \subset L^2$ which vanishes for $x>0$ but not identically and with its Fourier transform $G(x)$ satisfying $|G(x)| = \phi(x)$. Finally we have

$$\int_{-\infty}^{\infty} |F^{(\nu)}(x)|^2 dx = \int_{-\infty}^{\infty} |G(x)|^2 x^{2\nu} dx$$

$$= \int_{-\infty}^{\infty} \phi(x)^2 x^{2\nu} dx \leqq \int_{-\infty}^{\infty} [10(1+x^2)]^{-1}\left(\frac{x^{2\nu}}{A_\nu^2}\right)^{-1} x^{2\nu} dx \leqq A_\nu^2.$$

Thus the divergence of the integral (1) is certainly necessary for the quasi-analyticity of C_A.

Suppose now that $f(x)$ vanishes with all its derivatives at $x=0$, but does not vanish identically. We are to show that the integral (1) converges. Let $F(x)$ be identical with $f(x)$ for negative x and vanish identically for positive x, and let $G(x)$ be its Fourier transform. Then, assuming, as we may, that, in formula (*), $B=1$, we have

$$A_\nu^2 \geqq \int_{-\infty}^{\infty} |f^{(\nu)}(x)|^2 dx \geqq \int_{-\infty}^{\infty} |F^{(\nu)}(x)|^2 dx \geqq \int_{-\infty}^{\infty} |G(x)|^2 x^{2\nu} dx.$$

It follows that

$$\log\left(\sum_{\nu=0}^{\infty} \frac{r^{2\nu}}{A_\nu^2}\right) \leqq \log\left(\sum_{\nu=0}^{\infty}\left[\int_{-\infty}^{\infty} |G(x)|^2 \left(\frac{x}{r}\right)^{2\nu} dx\right]^{-1}\right)$$

$$\leqq \log\left(\sum_{\nu=0}^{\infty}\left[\int_{2r}^{2r+1} |G(x)|^2 \left(\frac{x}{r}\right)^{2\nu} dx\right]^{-1}\right)$$

$$\leqq \log\left(2\left[\int_{2r}^{2r+1} |G(x)|^2 dx\right]^{-1}\right) \leqq 2\int_{2r}^{2r+1} |\log(2^{-1/2}|G(x)|)|\,dx.$$

Hence

$$\int_1^{\infty} \log\left(\sum_{\nu=0}^{\infty}\frac{r^{2\nu}}{A_\nu^2}\right)\frac{dr}{r^2} \leqq 2\int_1^{\infty}\frac{dr}{r^2}\int_{2r}^{2r+1} |\log(2^{-1/2}|G(x)|)|\,dx$$

$$\leqq 2\int_2^{\infty} |\log(2^{-1/2}|G(x)|)|\,dx \int_{x/2-1/2}^{x/2}\frac{dr}{r^2}$$

$$\leqq 20\int_2^{\infty} |\log(2^{-1/2}|G(x)|)|\,x^{-2}dx < \infty.$$

Thus the divergence of (1) is also sufficient for quasi-analyticity.

II. On conjugate functions

1. We prove the following theorem.

THEOREM. *Let $f(\theta)$ be an odd function defined in $(-\pi, \pi)$ and suppose that it is absolutely integrable and non-decreasing in this range. Let $\bar{f}(\theta)$ be the conjugate function. Then $\bar{f}(\theta) \subset L$.*

We may assume, without loss of generality, that, in addition to the above properties, $f(\theta)$ satisfies the condition of assuming only integral values, for it differs by at most 1 from a function having all these properties. Then the function conjugate to this step function will differ from $\bar{f}(\theta)$ by a function which is certainly of class L^2.

2. Let us invert the circle $|z|<1$ into the half plane $\Im Z>0$, so that the point $e^{i\theta}$ inverts into $X(\theta)=\tan \theta/2$. Suppose that $f(\theta)$ inverts into $n(X)$. Then $n(X)$ is also non-decreasing in $(-\infty, \infty)$, and

$$\int_{-\pi}^{\pi} |f(\theta)|\, d\theta = 2\int_{-\infty}^{\infty} \frac{|n(X)|}{1+X^2}\, dX. \tag{1}$$

Let the points $\pm\lambda_\nu$, $\lambda_1 \leqq \lambda_2 \leqq \cdots$, be those at which $n(X)$ increases by 1 (if $n(X)$ increases by more than 1 then λ_n is taken an appropriate number of times). We may assume without loss of generality that λ_n is never zero, or, what is the same thing, that $n(X)$ is zero in the neighborhood of the origin. Then the convergence of the second integral (1) implies that of the series

$$\sum_{n=1}^{\infty} \lambda_n^{-1}. \tag{2}$$

3. Now consider the branch of the function

$$\frac{i}{\pi} \sum_{\nu=1}^{\infty} \log\left(1 - \frac{Z^2}{\lambda_\nu^2}\right) \tag{3}$$

which takes the value 0 at the origin, and is regular in the half plane $\Im Z \geqq 0$, $Z \neq \pm\lambda_\nu$ (the function (3) certainly exists in virtue of (2)). We observe that the real part of the function (3) coincides with $n(X)$ on the real axis, and is indeterminate at the points λ_ν. Further it may easily be seen that the imaginary part of the function (3) on the real axis differs only by a constant

$$Ci = -\pi i \int_{-\infty}^{\infty} \frac{\sum_{\nu=1}^{\infty} \log\left(1 - \dfrac{X^2}{\lambda_\nu^2}\right)}{\pi(1+X^2)}\, dX$$

since

$$\int_{-\infty}^{\infty}\left\{\Re\left[\frac{1}{\pi}\sum_{\nu=1}^{\infty}\log\left(1-\frac{X^2}{\lambda_\nu^2}\right)\right]+C\right\}\frac{2dX}{1+X^2}=\int_{-\pi}^{\pi}\bar{f}(\theta)d\theta=0,$$

from the transform to the conjugate $\bar{f}(\theta)$ of $f(\theta)$. Thus it is sufficient to establish the existence of the integral

$$\frac{\pi}{2}\int_{-\pi}^{\pi}|\bar{f}(\theta)|\,d\theta-\pi^2C=\int_{-\infty}^{\infty}\left|\Re\left\{\sum_{\nu=1}^{\infty}\log\left(1-\frac{X^2}{\lambda_\nu^2}\right)\right\}\right|\frac{dX}{1+X^2}$$

$$\leqq\int_{-\infty}^{\infty}\left|\Re\left\{\sum_{\nu=1}^{\infty}\log\left(1-\frac{X^2}{\lambda_\nu^2}\right)\right\}\right|\frac{dX}{X^2}\cdot \tag{4}$$

The integral (4) does not exceed

$$\sum_{\nu=1}^{\infty}\int_{-\infty}^{\infty}\left|\log\left|1-\frac{X^2}{\lambda_\nu^2}\right|\right|\frac{dX}{X^2}\leqq\sum_{\nu=1}^{\infty}\lambda_\nu^{-1}\int_{-\infty}^{\infty}|\log|1-X^2||\frac{dX}{X^2}<\infty,$$

in virtue of (2), and our theorem is proved.

4. We observe finally that the restriction that $f(\theta)$ should be an odd function is an essential one. For suppose that the theorem were true without this restriction. Let $f(\theta)$ be a function which is positive and which increases in $(-\pi, \pi)$, satisfying the two conditions

$$\int_{-\pi}^{\pi}f(\theta)d\theta<\infty,\qquad\int_{-\pi}^{\pi}f(\theta)\log^+f(\theta)d\theta=\infty.$$

Then, on the assumption that the theorem is true in the extended form, we should have

$$\int_{-\pi}^{\pi}|\bar{f}(\theta)|\,d\theta<\infty.$$

But, by a theorem of M. Riesz,* the last inequality, together with the condition $f(\theta)>0$, implies that

$$\int_{-\pi}^{\pi}f(\theta)\log^+f(\theta)d\theta<\infty,$$

giving a contradiction.

* Shortly to be published in the Journal of the London Mathematical Society.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY,
CAMBRIDGE, MASS.

PFAFFIAN SYSTEMS OF SPECIES ONE*

BY

JOSEPH MILLER THOMAS

This paper begins a study of the minimum number of differentials in terms of which a Pfaffian system can be expressed, together with the related subjects of reduced and canonical forms for such systems. The particular case treated here is that in which the minimum number of differentials exceeds the number of equations by unity. Such a system is said to be of *species one*. After passive (completely integrable) systems it is the simplest type. It is characterized by the following property: the adjunction of a single (suitably chosen) equation to it gives a passive system.

It is shown how to reduce any system of species one to a form involving the minimum number of differentials.

A system is called *nested* if the above reduction can be effected simultaneously for it and all of its derived systems. It is shown how to reduce any nested system of species one to a canonical form. It results that any such system can be written as the sum of a set of *special* systems, and that it is characterized by a finite set of arithmetical invariants. There is also given a necessary and sufficient condition for the existence of a nested system of species one having a given set of integers as invariants.

A further study of systems of species one in relation to their derived systems requires a theory of systems of higher species, which it is hoped to develop in a subsequent article.

The methods employed are largely those developed by Cartan and expounded by Goursat, with whose book† we suppose the reader is familiar.

1. **Generalities on the species of a Pfaffian system.** A *Pfaffian* is a form linear in the differentials of a finite set of variables, the coefficients being analytic functions of the variables. A *Pfaffian system* is obtained by equating to zero a linearly independent (and therefore finite) set of Pfaffians.

The variables are denoted by $x^1, x^2, \cdots, x^n$.

The *class* of a system is the minimum number of variables in terms of which a system equivalent to it can be expressed. A system comprising a single equation is always of odd class $2\sigma+1$. The minimum number of differentials which can appear in an equivalent equation is known to be $\sigma+1$,

* Presented to the Society, December 30, 1931; received by the editors January 20, 1932.

† Goursat, E., *Leçons sur le Problème de Pfaff*, Paris, 1922.

and is consequently determined by the class. This is not true in general for systems comprising more than one equation.

Consider the family of varieties

$$f^1(x) = \text{const.}, \cdots, f^k(x) = \text{const.}, \tag{1.1}$$

one of which passes through every point of a region. The rank of the matrix

$$\left\|\frac{\partial f}{\partial x}\right\|$$

is assumed to be k. If the above family is integral for the Pfaffian system

$$\omega^1 = 0, \cdots, \omega^r = 0, \tag{1.2}$$

the ω's must be linear homogeneous combinations of the df's; and conversely. Hence, *the minimum number of f's defining a family of integral varieties one of which passes through every point of a region is the same as the minimum number of differentials in terms of which an equivalent system can be expressed.*

Let the minimum number of differentials for system (1.2) be denoted by $r+\sigma$. The non-negative integer σ so defined will be called the species of the system.* The justification for this name is that the species enables us to differentiate between systems not differing in class or genus. Thus for every system of two equations of class five, expressed in characteristic variables, the genus, as defined by Cartan, is unity, whereas the species may be either one or two.

A function which is not identically equal to a constant and whose constancy is implied by (1.2) is called an *integral* of that system. Sometimes by ellipsis this term is also used to designate an *integral variety*, which we shall always designate by its full name, reserving the name integral for the concept to which it was attached by Poincaré.

The maximum number of independent integrals which the system (1.2) can possess is r and is attained only when the system is *passive*,† i.e., when it is equivalent to a system of the form

* The notion of species is inherent in the paper of E. von Weber, *Theorie der Systeme Pfaff'scher Gleichungen*, Mathematische Annalen, vol. 55 (1902), pp. 386–440. Although he does not mention the invariance of the minimum number of differentials, von Weber considers the possibility of reducing a given system to a form containing a specified number of differentials. In the terminology of the present paper, he gives necessary and sufficient conditions that the species do not exceed a specified value for systems whose *Stufe* (=class minus number of equations) does not exceed six.

† The usual term is *completely integrable* or *in involution*. We prefer to extend Riquier's terminology to Pfaffian systems. The justification of this lies in the fact that a Pfaffian system passive in the sense just defined is equivalent to a system of partial differential equations which is passive in Riquier's sense.

In the general theory of partial differential equations, the term passive is applicable to systems for which an existence theorem has not been proved and for which the name completely integrable would be at least temporarily a misnomer.

$$(1.3) \qquad dx^1 = 0, \cdots, dx^r = 0.$$

Passive systems are therefore systems of r equations whose class is r. Their species is zero.

It is known that the differentials appearing in a non-passive system can always be made fewer than the variables by at least unity. A consequence of this is that no system of r equations can be of class $r+1$, and that *the maximum value of the species for a system of class* p *is* $p-r-1$.

2. **The maximum system of species zero implied by a given system is determined by a maximum set of independent integrals.** The latter can be found from the last derived system of (1.2). A more direct method of finding it is furnished by

THEOREM 1. *The integrals of the Pfaffian system*

$$\omega^1 = 0, \cdots, \omega^r = 0$$

are the non-trivial solutions of the linear homogeneous partial differential equation of the first order

$$(2.1) \qquad \omega^1 \cdots \omega^r df = 0.$$

The proof is almost immediate. Since the equations of the Pfaffian system are independent, we have

$$(2.2) \qquad \omega^1 \cdots \omega^r \neq 0.$$

Equation (2.1) is then a necessary and sufficient condition* for the existence of multipliers λ such that

$$(2.3) \qquad df = \lambda_1\omega^1 + \cdots + \lambda_r\omega^r.$$

This proves the theorem.

Suppose f^1 is a non-trivial solution of (2.1). If we have also

$$(2.4) \qquad \omega^2 \cdots \omega^r df^1 = 0,$$

from (2.3) and (2.2) we deduce that the λ_1 in (2.3) is zero. If all r expressions of which the left member of (2.4) is the prototype were zero, equation (2.3) would show that f^1 is a constant, contrary to hypothesis. It is only a matter of notation, therefore, to assume that f^1 does not satisfy (2.4).

When the notation has been properly adjusted, system (1.2) is accordingly equivalent to

$$(2.5) \qquad df^1 = 0, \omega^2 = 0, \cdots, \omega^r = 0,$$

---

* Cartan, E., Bulletin de la Société Mathématique de France, vol. 29 (1901), p. 250.

and the integrals of (1.2) and (2.5) are the same. But the latter are by Theorem 1 the solutions of

$$\omega^2 \cdots \omega^r df^1 df = 0, \tag{2.6}$$

where f^1 is given and f is to be determined.

If (1.2) has more than one integral in its complete set, system (2.6) will have a solution f^2 for which $df^1 df^2 \neq 0$. From the discussion given above in the case of f^1, we know that adjusting the notation will make

$$\omega^3 \cdots \omega^r df^1 df^2 \neq 0,$$

and (1.2) is equivalent to

$$df^1 = 0,\ df^2 = 0, \omega^3 = 0, \cdots, \omega^r = 0.$$

If another integral f^3 exists, it is a solution of

$$\omega^3 \cdots \omega^r df^1 df^2 df = 0.$$

We continue until we reach an equation which has only a trivial solution. In this way, the complete set of q integrals is found. At the same time, we have a method of writing (1.2) in the form

$$dx^1 = 0, \cdots, dx^q = 0, \omega^{q+1} = 0, \cdots, \omega^r = 0, \tag{2.7}$$

which puts in evidence its maximum system of species zero written in the form (1.3). If $q=r$, the system is passive and the method reduces it to the form (1.3).

3. **Determination of the species.** The system

$$\omega^1 = 0, \cdots, \omega^r = 0, df^{r+1} = 0, \cdots, df^{r+k} = 0, \tag{3.1}$$

where

$$\omega^1 \cdots \omega^r df^{r+1} \cdots df^{r+k} \neq 0, \tag{3.2}$$

is passive if and only if (1.2) can be expressed in terms of $r+k$ differentials. The problem of finding a minimum set of differentials is therefore equivalent to that of finding a minimum passive Pfaffian system which implies the given one.

If we put

$$\omega^1 \cdots \omega^r \omega'^\alpha = \Omega^\alpha, \tag{3.3}$$

the conditions of passivity* of (3.1) are

$$\Omega^\alpha df^{r+1} \cdots df^{r+k} = 0. \tag{3.4}$$

* Cartan, E., *Leçons sur les Invariants Intégraux*, Paris, 1922, p. 101.

To determine the species, therefore, we form (3.1) and (3.2) for $k=0, 1, \cdots$ until the first consistent system is reached. The k at this stage is the species.

4. **Reduced form for systems of species one.** A system (1.2) is in *reduced form* if it involves the minimum number of differentials and is solved for r of them. The set of variables whose differentials appear in a reduced form will likewise be described by the adjective reduced.

Consider a system of species one expressed in the minimum number of differentials. Since the system is not passive, there are at least two of the $r+1$ differentials whose vanishing is not implied by the system. Call them dx^r and dx^{r+1}. Algebraically considered, (1.2) is a linear and homogeneous system of rank r in $r+1$ unknowns dx, and the unknown dx^{r+1} in particular can be chosen arbitrarily. Hence (1.2) can be written in the form

$$dx^1 - A^1dx^{r+1} = 0, \cdots, dx^r - A^rdx^{r+1} = 0. \tag{4.1}$$

If (1.2) is written in the form (2.7), the system

$$\omega^{q+1} = 0, \cdots, \omega^r = 0 \tag{4.2}$$

can be put in the form (4.1) because q of the $r+1$ differentials can be eliminated by means of $dx^1 = \cdots = dx^q = 0$. Hence (1.2) can be written as

$$dx^1 = 0, \cdots, dx^q = 0, dx^{q+1} - A^{q+1}dx^{r+1} = 0, \cdots, dx^r - A^rdx^{r+1} = 0, \tag{4.3}$$

where the A's are all different from zero, and *the maximum system of species zero is*

$$dx^1 = 0, \cdots, dx^q = 0. \tag{4.4}$$

5. **Reduced variables for systems of species one.** If $k=1$, system (3.4) is

$$\omega^1 \cdots \omega^r\omega'^{\alpha}df = 0 \qquad (\alpha = 1, 2, \cdots, r), \tag{5.1}$$

and is equivalent to a system of linear homogeneous partial differential equations of the first order in a single unknown function f. From §3, if (5.1) has a solution not satisfying (2.1), the system (1.2) is of species one or zero. Hence we have

THEOREM 2. *A non-passive Pfaffian system* (1.2) *is of species one if and only if the auxiliary system* (5.1) *has a solution which is not an integral of* (1.2).

Let f^{r+1} be a solution of (5.1) which does not satisfy (2.1). The system

$$\omega^1 = 0, \cdots, \omega^r = 0, df^{r+1} = 0 \tag{5.2}$$

is passive. Its integrals are a set of reduced variables for (1.2).

It was seen in §4 that if $f^1, \cdots, f^{r+1}$ constitute a set of reduced variables

any one of them which is not an integral of the system can be made to play the role of x^{r+1} in the discussion of reduced form. Adjoining $dx^{r+1}=0$ to a reduced form of the equations obviously gives a passive system of $r+1$ equations. Hence any reduced variable which is not an integral can play the role of f^{r+1} in (5.2) and will therefore satisfy (5.1). Since the integrals obviously satisfy (5.1), *all the reduced variables satisfy* (5.1).

If (5.1) has only $r+1$ independent solutions, any $r+1$ independent solutions will be a set of reduced variables because any set of reduced variables will be functions of those $r+1$ solutions, and the differentials of the reduced variables will be linear homogeneous combinations of the differentials of the $r+1$ solutions, so that the system can be expressed in terms of the latter set of differentials. Hence we have

THEOREM 3. *For a system of species one the auxiliary system* (5.1) *has at least* $r+1$ *independent solutions. If it has exactly* $r+1$, *any set of* $r+1$ *independent solutions is a set of reduced variables.*

An example is

$$dx^1 - x^2dx^5 = 0,\ dx^2 - x^1dx^3 - x^4dx^5 = 0. \tag{5.3}$$

The auxiliary system is

$$x^1\frac{\partial f}{\partial x^2} + \frac{\partial f}{\partial x^3} = 0,\quad \frac{\partial f}{\partial x^4} = 0.$$

It has just three solutions: $x^1, x^5, x^2-x^1x^3$. By the use of them, system (5.3) can be written in the reduced form

$$dx^1 - x^2dx^5 = 0,\ d(x^2 - x^1x^3) - (x^4 - x^2x^3)dx^5 = 0.$$

The auxiliary system can, however, have more than $r+1$ solutions. An example is furnished by the system

$$dx^1 - x^2dx^3 = 0, \tag{5.4}$$

whose auxiliary system has the $r+2$ solutions x^1, x^2, x^3.

Since the forms (3.1) are of degree $r+2$, except when they are zero, they cannot have more than $r+2$ linear factors. Consequently, *the number of independent solutions of the auxiliary system never exceeds* $r+2$ *unless* (1.2) *is passive.* When this maximum number of solutions is attained is stated by

THEOREM 4. *The auxiliary system for a system of species one has* $r+2$ *independent solutions if and only if the class is* $r+2$.

If the class is $r+2$, let $x^1, \cdots, x^{r+2}$ be characteristic variables. When the forms (3.1) are expressed in terms of these x's, they become

$$\Omega^\alpha = B^\alpha dx^1 \cdots dx^{r+2} \qquad (\alpha = 1, 2, \cdots, r), \tag{5.5}$$

because they are of degree $r+2$ in $r+2$ differentials. Hence the auxiliary system (5.1) has the $r+2$ x's for solutions.

Conversely, let $x^1, \cdots, x^{r+2}$ be independent solutions of (5.1). Equations (5.5) then hold. From them we find the characteristic system* of (1.2) to be

$$dx^1 = 0, \cdots, dx^{r+2} = 0.$$

Hence the class is $r+2$, and the theorem is proved.

When the auxiliary system has $r+2$ solutions, $r+1$ of them chosen at random do not necessarily form a set of reduced variables. Thus x^1 and x^3 do not form such a set for (5.4). When one solution f^{r+1}, other than an integral, has been found for the auxiliary system, the set of reduced variables is completed by finding the integrals of (5.2). This can be accomplished by the method developed in §2; for example, by solving

$$\omega^1 \cdots \omega^r df^{r+1} df = 0. \tag{5.6}$$

For the example (5.4) with $f^{r+1} = x^3$, system (5.6) is

$$x^3 \frac{\partial f}{\partial x^1} + \frac{\partial f}{\partial x^2} = 0,$$

of which a solution is $x^1 - x^2x^3$. Hence a set of reduced variables containing the variable x^3 is x^3, $x^1 - x^2x^3$, and the equation can be written

$$d(x^1 - x^2x^3) + x^2dx^3 = 0.$$

To put a given system in reduced form, write it first in the form (2.7), and then consider the auxiliary system for (4.2). When one solution of the latter has been found, the set of reduced variables can be completed as indicated above for (5.6).

6. **Properties of the reduced form for systems of species one.** For a system in the form (4.3) the derived forms ω', with $dx^1, \cdots, dx^r$ eliminated by means of (4.3), are

$$\omega'^1 = 0, \cdots, \omega'^q = 0, \quad \omega'^{q+1} = -\partial A^{q+1} dx^{r+1}, \cdots, \omega'^r = -\partial A^r dx^{r+1}, \tag{6.1}$$

where ∂ denotes the differential formed on the assumption

$$x^1 = \text{const.}, \cdots, x^{r+1} = \text{const.} \tag{6.2}$$

Now one of the ∂A's must be different from zero. Otherwise all the derived

* Cartan, loc. cit. in §3 of this paper.

forms would be zero and the system would be passive. Hence the characteristic system contains the equations

$$dx^1 = 0, \cdots, dx^{r+1} = 0,\ \partial A^{q+1} = 0, \cdots, \partial A^r = 0.$$

Its rank, the class p of the system, is accordingly the number of independent variables in the set

$$x^1, \cdots, x^{r+1};\ A^{q+1}, \cdots, A^r. \tag{6.3}$$

The rank of the matrix

$$\left\| \begin{matrix} \dfrac{\partial A^{q+1}}{\partial x^{r+2}}, & \cdots, & \dfrac{\partial A^{q+1}}{\partial x^n} \\ \cdots, & \cdots, & \cdots \\ \dfrac{\partial A^r}{\partial x^{r+2}}, & \cdots, & \dfrac{\partial A^r}{\partial x^n} \end{matrix} \right\| \tag{6.4}$$

is therefore $p-r-1$.

The number of linearly independent solutions of

$$\lambda_{q+1}\partial A^{q+1} + \cdots + \lambda_r \partial A^r = 0$$

is the number of unknowns minus the rank of (6.4). This number, increased by q corresponding to equations (4.4), gives the number of equations r^1 in the first derived system:

$$r^1 = 2r + 1 - p. \tag{6.5}$$

7. **The genus of a Pfaffian system.** The genus of a Pfaffian system, as defined by Cartan,* satisfies the equation

$$n - \gamma = r + s_1 + s_2 + \cdots, \tag{7.1}$$

where the sum on the right extends over all the characters s. The number γ is the maximum dimensionality of a non-singular integral variety.

In studying Pfaffian systems, however, it is customary to consider changes of variables which do not preserve dimensionality, i.e., the number of dimensions of the representative space may change. Hence γ is not an invariant with respect to the transformations considered. This leads to some confusion, which is perhaps only increased by defining a "true" genus.†

On the other hand, the non-zero characters are invariant under the trans-

* Cartan, E., *Sur l'intégration des systèmes d'équations aux différentielles totales*, Annales Scientifiques de l'Ecole Normale Supérieure, (3), vol. 18 (1901), p. 262.

† Cf. Goursat, p. 362; Cartan, p. 291.

formations in question. Consequently, the left member of (7.1) is invariant. Its significance is ***the minimum number of independent relations defining a non-singular integral variety***. Its least value is r. We shall write

$$g = n - \gamma - r = s_1 + s_2 + \cdots . \tag{7.2}$$

Because of its invariance, we shall employ g rather than γ, and in order not to multiply names needlessly, we shall call g the genus rather than γ. The least value of g is zero and occurs for a passive system. For a non-passive system, the value of γ computed in characteristic variables is at least one. Hence, *the genus of a non-passive system satisfies the inequalities*

$$0 < g \leqq p - r - 1. \tag{7.3}$$

An integral variety defined by fewer than $g+r$ relations is necessarily singular.

8. **The genus of systems of species one.** The system* whose rank is the character $s = s_1$ of system (4.3) is readily found by use of (6.1) to be

$$\begin{aligned} &\left(\frac{\partial A^i}{\partial x^{r+2}}dx^{r+2} + \cdots + \frac{\partial A^i}{\partial x^p}dx^p\right)\delta x^{r+1} \\ &\quad - dx^{r+1}\left(\frac{\partial A^i}{\partial x^{r+2}}\delta x^{r+2} + \cdots + \frac{\partial A^i}{\partial x^p}\delta x^p\right) = 0 \qquad (i = q+1, \cdots, r), \end{aligned} \tag{8.1}$$

where the variables are characteristic (i.e., $n = p$), the dx's are given, and the ∂x's are the unknowns.

If $dx^{r+1} \neq 0$, multiplying the second, third, $\cdots$, $(p-r)$th columns of the matrix of (8.1) by dx^{r+2}/dx^{r+1}, dx^{r+3}/dx^{r+1}, $\cdots$, dx^p/dx^{r+1}, respectively, and adding to the first reduces it to a column of zeros. Hence the rank of (8.1) is that of (6.4), namely, $p - r - 1$. Thus the value of the character is

$$s = p - r - 1, \tag{8.2}$$

a formula which of course only applies to non-passive systems. Since the genus is at least equal to s, inequality (7.3) gives

$$s = g, \tag{8.3}$$

a result which we state as

THEOREM 5. *For any system of species one the genus and character are equal.*

Because of (8.3) equation (6.5) can be written

$$s = r - r^1. \tag{8.4}$$

---

* Goursat, p. 290.

9. **Systems of species one whose first derived system has species not exceeding one.** If the first r^1 equations of a system S written in reduced form (4.1) constitute its derived system S^1, the auxiliary systems (5.1) for S and S^1 have in common the solution x^{r+1}, which is not an integral of S or S^1.

Conversely, if the auxiliary systems of S and S^1 have in common a solution x^{r+1} which is not an integral, let a set of reduced variables containing x^{r+1} be found for S^1. By means of it S^1 is put in the form

$$(9.1) \qquad S^1:\ dx^1 - A^1dx^{r+1} = 0, \cdots, dx^{r^1} - A^{r^1}dx^{r+1} = 0.$$

Clearly $x^1, \cdots, x^{r^1}$ are integrals of the passive system obtained by adjoining $dx^{r+1}=0$ to S. Since any set of $r+1$ independent integrals of that passive system is a set of reduced variables for S, a set of reduced variables for S containing $x^1, \cdots, x^{r^1}, x^{r+1}$ can be found, whereby the equations Σ which it is necessary to adjoin to S^1 in order to get S can be given the form

$$(9.2) \qquad \Sigma:\ dx^{r^1+1} - A^{r^1+1}dx^{r+1} = 0, \cdots, dx^r - A^r dx^{r+1} = 0.$$

The conditions that S^1 be the derived system of S can be written

$$(9.3) \qquad \omega^1 \cdots \omega^r \omega'^\alpha = 0 \qquad (\alpha = 1, 2, \cdots, r^1).$$

Substitution from (9.1) and (9.2) gives

$$(9.4) \qquad dx^1 \cdots dx^{r+1} dA^\alpha = 0 \qquad (\alpha = 1, 2, \cdots, r^1).$$

Consequently $A^1, \cdots, A^{r^1}$ are functions of $x^1, \cdots, x^{r+1}$ alone:

$$(9.5) \qquad A^1 = A^1(x^1, \cdots, x^{r+1}), \cdots, A^{r^1} = A^{r^1}(x^1, \cdots, x^{r+1}).$$

Because of the statement just preceding (6.3), relations (9.5) show that the number of independent variables in the set

$$(9.6) \qquad x^1, \cdots, x^{r+1}, A^{r^1+1}, \cdots, A^r$$

is the class p. From (6.5) the class is $2r-r^1+1$. Since this is the number of variables (9.6), those variables form an independent set.

If S^1 is passive, any one of its reduced variables is an integral of S and therefore satisfies the auxiliary system for S.

If S^1 is not passive and its class is not r^1+2, by Theorems 3 and 4 any reduced variable for S^1 is a function of the $x^1, \cdots, x^{r^1}, x^{r+1}$ in (9.1) and therefore satisfies the auxiliary system of S.

If S^1 is of class r^1+2, one of the coefficients in (9.1) must involve one of the variables $x^{r^1+1}, \cdots, x^r$. Suppose one involves x^r. A change of variable will make the coefficient in question x^r. At the same time the form of the last equation of Σ can be preserved by using the equations of S to eliminate the

excessive differentials. Direct calculation then shows that x^r is a characteristic variable and consequently a reduced variable for S^1; and from the form of (9.1), (9.2) x^r is reduced for S.

In every case, therefore, *if a variable which is not an integral is reduced for both S^1 and S, every variable reduced for S^1 is reduced for S also.*

As a result of the developments in this section we have

THEOREM 6. *Given a system S of species one whose first derived system S^1 has species not exceeding one. Any reduced form of S^1 can be made the first r^1 equations of a reduced form of S if and only if the auxiliary systems of S and S^1 have in common a solution which is not an integral.*

Simultaneous reduction is not always possible. The system

$$
\begin{gathered}
dx^1 - x^2dx^4 = 0, \\
dx^2 - x^5dx^4 = 0,\ dx^2 - x^6dx^3 - x^4dx^5 = 0
\end{gathered} \tag{9.7}
$$

is of species one. Its derived system is the first equation and is also of species one. But S and S^1 cannot be thrown simultaneously into reduced form because their auxiliary systems have no solution in common.

10. **Nested systems of species one. Canonical form.** If the auxiliary systems (5.1) formed for a system and all its derived systems, except the last, have in common a solution which is not an integral, the system will be called a *nested system of species one.* All the derived systems, except the last, are systems of the same sort.

By a *canonical form* of a Pfaffian system we mean an equivalent system in which the variables are all independent and each equation is in the canonical form for a single equation.*

We next prove

THEOREM 7. *Every nested system of species one can be reduced to a canonical form having the following properties:*

(i) *The system S is in reduced form.*

(ii) *The first derived system S^1 is obtained by omitting the last $r-r^1$ equations of S, the second derived system S^2 by omitting the last r^1-r^2 equations of S^1, etc.*

(iii) *Every variable occurs at most once as a coefficient and at most once as a differential with the exception of the variable whose differential occurs in all the second terms.*

(iv) *Any variable which occurs as a coefficient but not as a differential in S^i occurs as a differential in S^{i-1}.*

(v) *The only variables which do not occur as coefficients are those whose dif-*

* See Goursat, p. 58, for a definition of the latter.

ferentials occur in the last derived system and the variable whose differential occurs in all the second terms.

(vi) *The only variables whose differentials do not occur at all are the coefficients in the equations which are in S but not in S^1.*

It is of course clear that every derived system will be in a canonical form having the same properties.

The theorem is immediate for $r=1$.

We now assume the theorem for $r-1$ equations and proceed by induction.

The given system S is not passive, but since we are to apply induction and one of the derived systems may be passive, it is necessary to consider the passive case and to remark that a passive system can obviously be thrown into a canonical form having all the enumerated properties.

Since S is not passive, the number of equations in its derived system does not exceed $r-1$. The first derived system can therefore be supposed written in canonical form as the first r^1 equations of S:

$$\begin{gathered} dx^1 = 0, \cdots, dx^q = 0,\ dx^{q+1} - x^\alpha dx^{r+1} = 0, \cdots, \\ dx^{r^1} - x^\beta dx^{r+1} = 0. \end{gathered} \tag{10.1}$$

Theorem 6 can be applied to show that the remaining equations of S can be simultaneously given the form

$$dx^{r^1+1} - A^{r^1+1}dx^{r+1} = 0, \cdots, dx^r - A^r dx^{r+1} = 0. \tag{10.2}$$

Let the variables which occur as coefficients but not as differentials in S^1 be denoted by

$$\bar{x}^{r^1+1}, \cdots, \bar{x}^{r^1+s^1}. \tag{10.3}$$

Relations (9.5) give

$$\bar{x}^{r^1+1} = F^{r^1+1}(x^1, \cdots, x^{r+1}), \cdots, \bar{x}^{r^1+s^1} = F^{r^1+s^1}(x^1, \cdots, x^{r+1}). \tag{10.4}$$

It must be possible to solve (10.4) for s^1 of the variables

$$x^{r^1+1}, \cdots, x^r, \tag{10.5}$$

for otherwise their elimination would lead to a relation among

$$x^1, \cdots, x^{r^1}, \bar{x}^{r^1+1}, \cdots, \bar{x}^{r^1+s^1}, x^{r+1},$$

that is, among the variables in terms of which S^1 is expressed. This would contradict S^1's being in canonical form. In particular, we must have

$$s^1 \leqq r - r^1. \tag{10.6}$$

Hence by means of (10.4) we can express the differentials of s^1 of the variables (10.5), which it is merely a matter of notation to suppose are

$$x^{r^1+1}, \cdots, x^{r^1+s^1}, \tag{10.7}$$

as linear, homogeneous combinations of

$$dx^1, \cdots, dx^{r^1}, d\bar{x}^{r^1+1}, \cdots, d\bar{x}^{r^1+s^1}, dx^{r^1+s^1+1}, \cdots, dx^{r+1}.$$

By use of the equations of S^1 these become combinations of

$$d\bar{x}^{r^1+1}, \cdots, d\bar{x}^{r^1+s^1}, dx^{r^1+s^1+1}, \cdots, dx^{r+1}.$$

When these values have been substituted in (10.2), equations (10.2) can be put by solving into a similar form in which the differentials of the variables (10.7) are replaced by those of (10.3). So all x's occurring as coefficients in S^1 occur as differentials in the new form of S. Moreover, the A's in the new (10.2) when taken with the x's form a set of independent variables because of the result about the variables (9.6). The induction is therefore complete.

If the derived system is vacuous, the demonstration holds as given, but also follows immediately from the reduced form and the result about the variables (9.6).

It is clear that *any system written in the above canonical form is nested and of species one.*

The number s^1 appearing in above discussion and in (10.6) is the number of equations in S^1 but not in S^2, for it gives the number of ω''s which vanish by virtue of S but not by virtue of S^1. Hence we have

$$s^1 = r^1 - r^2. \tag{10.8}$$

In the same way we denote by s^i the number of variables which occur as coefficients but not as differentials in S^i. From (10.8) we have therefore

$$s^i = r^i - r^{i+1}. \tag{10.9}$$

From (8.4) it is seen that the s's are the first characters of the successive derived systems, s^0 being s. From (8.4) and (10.6) we get

$$s \geqq s^1 \geqq s^2 \geqq \cdots \geqq 0. \tag{10.10}$$

We find it convenient to define a set of integers t by means of the formulas

$$t^i = s^i - s^{i+1}. \tag{10.11}$$

These t's are the second differences of the numbers r. Even for a nested system of species one they do not necessarily satisfy inequalities like (10.10), but they are non-negative because the s's satisfy (10.10).

Let l be the least number such that the derived system S^l is passive or vacuous. Since

$$r^l = r^{l+1},$$

we have

$$s^i = t^i = 0 \qquad (i \geqq l). \tag{10.12}$$

As a result of the foregoing, the numbers s, t are all determined for a nested system of species one when the r's are given. Conversely, if t, $t^1, \cdots,$ t^{l-1} and r^l are given, the s's are determined by (10.11) and (10.12). The other r's are then determined by (10.9).

11. **A nested system of species one as a sum of special systems.** Consider a nested system of species one written in the canonical form of the preceding section. From it construct a set of systems in the following manner. Make $\omega^r=0$, which is the last equation in S, the last equation of one system. If x^r, whose differential is the first term of ω^r, is not a coefficient in S^1, the new system will consist of the single equation $\omega^r=0$ and will be called a system T_0. If x^r is the coefficient in an equation of S^1, we take that equation, say $\omega^{r^1}=0$, as the next to the last in the new system. If x^{r^1} is not a coefficient in S^2, we close the new system and call it a system T_1. If x^{r^1} is a coefficient in S^2, we take the corresponding equation, say $\omega^{r^2}=0$, as the third from the last in the new system. And so on. If the system is closed with an equation from S^i, it will be called a system T_i. It contains $i+1$ equations.

We next take the equation $\omega^{r-1}=0$, if it is not in S^1, and form in the above manner the system T which it determines. And so on until all the equations in $\Sigma=S-S^1$ are exhausted. When this stage is reached, all the equations of S, except those in its passive system, have been used. This is for the following reason. Because of the nature of the canonical form, the coefficient in an equation of S^i-S^{i+1} occurs as a differential in an equation of $S^{i-1}-S^i$; the coefficient in the latter occurs as a differential in $S^{i-2}-S^{i-1}$; and so on until $S-S^1$ is reached.

The passive system is to be taken as a separate system. For it there is already the notation S^l.

The systems T are nested, of species one and in canonical form. In addition, *each derived system is obtained from the preceding by omitting its last equation*, and the class exceeds the number of equations by two. Consequently, each system T is special and of the type first reduced to canonical form by von Weber.* A passive system is also special. Therefore we have

* See Goursat, pp. 321–8.

THEOREM 8. *Every nested system of species one is equivalent to a set of special systems no two of which have an equation in common.*

Let t^i be the number of special systems T_i. There is one and only one equation in $S-S^1$ corresponding to each system T. Hence the total number of equations in $S-S^1$ is the same as the total number of systems T, a fact which is stated by the equation

$$r - r^1 = t^0 + t^1 + \cdots + t^{l-1}.$$

In the same way, there are just as many equations in S^1-S^2 as there are systems T with index at least unity. Hence we get the sequence of relations

$$\begin{aligned} r^1 - r^2 &= t^1 + \cdots + t^{l-1}, \\ r^2 - r^3 &= t^2 + \cdots + t^{l-1}, \\ \cdots \quad & \quad \cdots \quad \cdots \\ r^{l-1} - r^l &= t^{l-1}. \end{aligned}$$

From these and (10.9) it readily follows that *the t's are the numbers defined by* (10.11).

There is only one symbol common to two of the partial systems, namely, dx^{r+1}, and it plays the same role throughout. It is clear, therefore, that a nested system of species one can be broken up by the above process into a sum of special systems in only one way and that two nested systems of species one are equivalent if and only if their sets of special systems can be transformed one into the other.

From the canonical form of a special system whose class is two more than the number of equations and which has no integrals it follows immediately that *two such systems are equivalent if and only if they consist of the same number of equations.** Therefore two nested systems of species one are equivalent if and only if they have the same number of integrals and the same sequence of numbers t. From the results at the end of §10 this amounts to having the same number of integrals and the same sequence of first characters; or to having the same sequence of numbers r.

THEOREM 9. *A nested system of species one is completely characterized by a finite set of arithmetic invariants, which can be taken as the number of equations in the successive derived systems.*

If $(r, r^1, \cdots, r^l)$ is defined as the *symbol* of a Pfaffian system, *a nested system of species one is completely characterized by its symbol.*

The following question arises: what conditions must the non-negative

* Goursat, p. 328.

integers in $(r, r^1, \cdots, r^l)$ satisfy in order that it may be the symbol of a nested system of species one? In the first place, to have a meaning at all, it must not contain two equal r's because of the definition of l. In the second place, the characters formed as the first differences of the numbers in the symbol must be non-increasing and non-negative in accordance with (10.10); and from the fact that no two r's are equal they must therefore be positive. These two conditions are also sufficient. For the satisfaction of (10.10) insures that the t's will be non-negative. Corresponding to each t^i we write t^i special systems of $i+1$ equations and class $i+3$, each in canonical form with the dx^{r+1} in common and all the other variables distinct. To these we adjoin the differentials of r^l other variables equated to zero. The result is a nested system of species one having the specified symbol.

THEOREM 10. *There is a nested system of species one having* $(r, r^1, \cdots, r^l)$ *as symbol if and only if the first differences of the r's form a non-increasing sequence of positive integers.*

If the given system is special and of class $r+2$, the invariants have the following values:

$$l = r - q, \quad s^0 = s^1 = \cdots = s^{l-1} = 1, \quad t^0 = t^1 = \cdots = t^{l-2} = 0, \quad t^{l-1} = 1.$$

DUKE UNIVERSITY,
DURHAM, N. C.

FAMILIES OF GROUPS GENERATED BY TWO OPERATORS OF THE SAME ORDER*

BY

ABRAHAM SINKOV

I. Introduction

The present paper is an extension of a paper by W. E. Edington† entitled *On an infinite system of non-abelian groups of order* nm^{n-1}, in which it was shown that given any two numbers n and m, there exists a group of order nm^{n-1} generated by two operators of order n. Edington's proof requires the assumptions that all the operators of the form $S_1^{\alpha}S_2^{\alpha}$ are commutative and of the same order; it follows from the present treatment that the second is a consequence of the first and may be dispensed with. The results herein obtained exhibit a more general system of groups of order nm^{n-k}, where k is an arbitrary factor of n, and obtain some properties of these groups not considered by Edington.

II. A property of groups G generated by two operators of the same order

To begin with, the generating operations S_1 and S_2 are assumed to be of the same order n, so that $S_1^n = S_2^n = 1$. As yet no further restrictions are supposed. A third relation which defines the order of a particular combination of S_1 and S_2 will be introduced later on, in order to show how it actually arises.

Under these conditions, then, consider the totality of operators

$$S_1^{\alpha}S_2^{n-\alpha} \qquad (\alpha = 1, 2, \cdots, n-1).$$

This set of operators

$$A:\begin{cases} S_1S_2^{n-1} \\ S_1^2S_2^{n-2} \\ \cdot\ \cdot\ \cdot \\ \cdot\ \cdot\ \cdot \\ S_1^{n-1}S_2 \end{cases}$$

* Presented to the Society, March 25, 1932; received by the editors September 12, 1932. This paper was written under the guidance of Professor F. E. Johnston.

† Annals of Mathematics, vol. 25 (1923), p. 85.

generates a sub-group H of G. The transform of the generator $S_1^{\alpha}S_2^{n-\alpha}$ by S_1 can be written

$$(S_1^{\alpha-1}S_2^{n-\alpha+1})(S_1^{n-1}S_2)^{-1}$$

and is in H. When transformed by S_2 this same generator becomes

$$(S_1^{n-1}S_2)^{-1}(S_1^{\alpha-1}S_2^{n-\alpha+1}),$$

a result which is again in H. Consequently, H is invariant in G.

Evidently, the adjunction of either S_1 or S_2 to H generates G. It follows that the index of H under G cannot exceed n; it will equal n if and only if H involves no power of either S_1 or S_2. Such will be assumed to be the case. The order of G is then equal to n times the order of H.

It is now possible to replace the operators $S_1^{\alpha}S_2^{n-\alpha}$, which generate H, by an equivalent set in the sense that this new set generates the same group, and possesses the added advantage that all of the new generators are of the same order. In fact, they are the complete set of conjugates of $S_1S_2^{n-1}$ under S_2:

$$S_2^{q}(S_1S_2^{n-1})S_2^{-q} = (S_1^{q}S_2^{n-q})^{-1}(S_1^{q+1}S_2^{n-q-1}) \qquad (q = 0, 1, 2, \cdots, n-1).$$

This new set will be denoted by ξ:

$$\xi: \begin{cases} S_1S_2^{n-1} \\ (S_1S_2^{n-1})^{-1}(S_1^{2}S_2^{n-2}) \\ (S_1^{2}S_2^{n-2})^{-1}(S_1^{3}S_2^{n-3}) \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ (S_1^{n-1}S_2)^{-1}. \end{cases}$$

Observe that this set is obtainable from the preceding one by multiplying each operator of A in turn by the inverse of the one which precedes it, provided that the identity is supposed to precede A_1 and to follow A_{n-1}. Conversely A is obtainable from ξ because of the relation

$$\prod_{i=1}^{\alpha}\xi_i = S_1^{\alpha}S_2^{n-\alpha}$$

where ξ_i denotes the ith term in the set ξ. In particular

$$\prod_{i=1}^{n}\xi_i = 1.$$

The preceding results may now be summarized as follows:

THEOREM I. *The group G generated by two operators of the same order n contains an invariant sub-group H generated by n operators of the same order. The index of H under G is at most n.*

It is not amiss to remark at this point that the common order m of these new generators of H is quite arbitrary and may be assumed to be any integer whatever.

III. DEFINING RELATIONS OF A 3-PARAMETER FAMILY

It will now be assumed that the sub-group H is abelian and that it is possible to select q of the operators ξ which shall form a set of independent generators in the restricted sense. Its order will then be m^q.*

Suppose for a moment that n is composite and contains the factor k. Then $n = kx$, and the set of operators ξ can be divided up in order into x sub-sets of k each. Suppose further that it is possible for the k operators of one of these sub-sets to be dependent upon the remaining generators, which form an independent set in the restricted sense. Then, under such circumstances the most general sub-group H which can be obtained, i.e., the one of greatest order, will be of order $m^{(x-1)k}$. By analogy with the condition previously obtained on the ξ_i, the equation of connection will be chosen in the form

$$\prod_{i=0}^{x-1} \xi_{ik+\alpha} = 1 \qquad (\alpha = 1, 2, \cdots, k;\ n = kx).$$

Expressed in words this means that the product of the operations which are in corresponding positions in each sub-set is the identity.

Since

$$\xi_{a+1} = S_2^a S_1 S_2^{kx-a-1},$$

$$\prod_{i=0}^{x-1} \xi_{ik+\alpha} = 1 = (S_2^{\alpha-1} S_1 S_2^{kx-\alpha})(S_2^{k+\alpha-1} S_1 S_2^{kx-k-\alpha}) \cdots (S_2^{(\ -1)k+\alpha-1} S_1 S_2^{k-\alpha})$$
$$= S_2^{\alpha-1}(S_1 S_2^{k-1})^x S_2^{-\alpha+1},$$

whence

$$(S_1 S_2^{k-1})^x = 1.$$

In the particular case $k = 1$, this condition reduces to an identity.

Conversely, if $(S_1 S_2^{k-1})^x = 1$, it follows from a reversal of the preceding manipulation that

$$\prod_{i=0}^{x-1} \xi_{ik+\alpha} = 1 \qquad (\alpha = 1, 2, \cdots, k),$$

---

* Miller, Blichfeldt and Dickson, *Theory and Applications of Finite Groups*, p. 90.

and hence that k of the operators ξ are expressible in terms of the remaining ones.

It is now necessary to examine this condition a bit more closely. Suppose it holds for some particular factor k of n. Then it can be shown that it is also satisfied by every factor of n which is a divisor of k. For, suppose $k=rt$. Then of the rt continued products

$$\prod_{i=0}^{x-1}\xi_{irt+\alpha} = 1 \qquad (\alpha = 1, 2, \cdots, rt),$$

select the following r:

$$\prod_{i=0}^{x-1}\xi_{irt+\alpha} = 1,$$

$$\cdots\cdots\cdots\cdots$$

$$\prod_{i=0}^{x-1}\xi_{irt+\alpha+(r-1)t} = 1 \qquad (\alpha = 1, 2, \cdots, t).$$

Multiplying together all of the left hand members and rearranging the terms (which is permissible because of the commutativity of the ξ_i),

$$\prod_{i=0}^{rx-1}\xi_{it+\alpha} = 1 \qquad (\alpha = 1, 2, \cdots, t).$$

The condition therefore holds for t.

This leads to the conclusion that if for a given S_1 and S_2 two or more values k satisfy the relation $(S_1S_2^{k-1})^x=1$, and if it is possible to choose one among them such that all the others are divisors of it, then that one is the value to be used in determining the number of independent generators of H.

Moreover, the assumption that the correct k requires the remaining $(x-1)k$ operators to be independent prevents two numbers, where neither is a multiple of the other, from simultaneously satisfying the required condition. For, in that case, there arises the obvious contradiction that the order of H is given by two different powers of m.

If the relation holds for no $k>1$, then $k=1$ and $q=n-1$.

The number k having been determined in this way, it follows that the order of H is m^{n-k}. Since it has already been shown that the index of H under G in the most general case is n, then the corresponding order of G is nm^{n-k}.

There thus results the following theorem:

THEOREM II. *Two operators S_1 and S_2 of the same order n for which the set ξ is commutative generate a group G whose order is at most nm^{n-k} where m is the common order of the operators ξ and the number k is defined as the greatest factor of n satisfying the relation*

$$(S_1S_2^{k-1})^x = 1 \qquad (kx = n).$$

IV. Generating operations of G

The maximum group thus defined exists for every number nm^{n-k} and a pair of generating operations S_1 and S_2, of a fairly simple form, can be set up for the general case. Let

$$S_1 = (a_1a_2 \cdots a_n)(a_{n+1}a_{n+2} \cdots a_{2n}) \cdots (a_{(m-1)n+1}a_{(m-1)n+2} \cdots a_{mn}),$$
$$S_2 = (a_{p+1}a_{p+2} \cdots a_{p+n})(a_{n+p+1}a_{n+p+2} \cdots a_{2n+p}) \cdots (a_{(m-1)n+p+1} \cdots a_{mn}a_1a_2 \cdots a_p).$$

Then it can be shown that the operators $S_1^{\alpha}S_2^{n-\alpha}$ are all of order m and are commutative. Moreover, if k is the greatest common divisor of p and n, it is found that $n-k$ of the operators ξ form an independent set in the restricted sense. Hence there results the following theorem:

Theorem III. *The two substitutions given above generate a group G of order nm^{n-k} where k is the greatest common divisor of p and n.*

V. Properties of the groups G

The quotient group of H under G is cyclic and as a consequence every G is solvable.

In order to determine the central it is first necessary to determine the subgroup of it which is contained in H, i.e., the combinations of the ξ_i which are invariant in G.

To obtain this result requires a knowledge of how the ξ_i are transformed under S_1 and S_2. First, consider the transform of ξ_{i+1} by S_1. Since

$$\xi_{i+1} = S_2^iS_1S_2^{n-i-1},$$
$$S_1^{-1}\xi_{i+1}S_1 = (S_1^{n-1}S_2)(S_1^{i-1}S_2^{n-i+1})^{-1}(S_1^iS_2^{n-i})(S_1^{n-1}S_2)^{-1}.$$

In this last result, the two factors $(S_1^{n-1}S_2)^{-1}$ and $(S_1^{n-1}S_2)$ will be eliminated as a result of the commutativity of the $S_1^{\alpha}S_2^{n-\alpha}$. Hence

$$S_1^{-1}\xi_{i+1}S_1 = (S_1^{i-1}S_2^{n-i+1})^{-1}(S_1^iS_2^{n-i}) = \xi_i.$$

Similarly

$$S_2^{-1}\xi_{i+1}S_2 = \xi_i$$

and the ξ_i are thus seen to be transformed in the same way by S_1 and S_2. This property is also obtainable from the fact that S_1 and S_2 are contained in the same co-set of H. As a consequence of it, the operators of H which are invariant under S_1 are identical with those which are invariant in G.

It has already been seen that for a given k the operators ξ can be divided up into k different sets

$$\xi_{ik+\alpha} \qquad (\alpha = 1, 2, \cdots, k)$$

each containing x operators, and such that $x-1$ of the operators in each set are independent. Moreover, no operators in any one of these sets can be expressed in terms of any of the other sets. Hence, since ξ_i and ξ_{i-1} are in different sets if $k>1$, it follows that a combination of operators in any single set can be invariant in G, only if $k=1$.

Suppose then that such is the case and that

$$\xi_1^{a_1}\xi_2^{a_2}\cdots\xi_{n-1}^{a_{n-1}} \qquad (a_i < m)$$

is invariant in G. (ξ_n may be omitted from this combination since it is expressible in terms of the remaining $(n-1)$ ξ's.) This operator is equal to its transform under S_1 and consequently the following relation must hold:

$$\begin{aligned}\xi_1^{a_1}\xi_2^{a_2}\cdots\xi_{n-1}^{a_{n-1}} &= S_1^{-1}(\xi_1^{a_1}\xi_2^{a_2}\cdots\xi_{n-1}^{a_{n-1}})S_1\\ &= \xi_n^{a_1}\xi_1^{a_2}\cdots\xi_{n-2}^{a_{n-1}}.\end{aligned}$$

This leads to

$$\begin{aligned}\xi_{n-1}^{a_{n-1}} &= \xi_n^{a_1}\xi_1^{a_2-a_1}\xi_2^{a_3-a_2}\cdots\xi_{n-2}^{a_{n-1}-a_{n-2}}\\ &= (\xi_n^{a_1}\xi_1^{a_1}\cdots\xi_{n-2}^{a_1})(\xi_1^{a_2-2a_1}\xi_2^{a_3-a_2-a_1}\cdots\xi_{n-2}^{a_{n-1}-a_{n-2}-a_1}).\end{aligned}$$

As a consequence of the relation

$$\prod_{i=1}^{n}\xi_i = 1$$

the first quantity in parentheses is reducible at once to $(\xi_{n-1})^{-a_1}$. Hence

$$\xi_{n-1}^{a_1+a_{n-1}} = \xi_1^{a_2-2a_1}\xi_2^{a_3-a_2-a_1}\cdots\xi_{n-2}^{a_{n-1}-a_{n-2}-a_1}.$$

This equality involves only $n-1$ of the ξ's, all of which are independent of one another. Therefore both members must reduce to the identity, and as a result we have the following series of congruences:

$$\left.\begin{aligned}a_1 &\equiv -\,a_{n-1}\\ a_2 &\equiv 2a_1\\ a_3 &\equiv a_2 + a_1\\ &\cdots\cdots\\ a_{n-1} &\equiv a_{n-2} + a_1\end{aligned}\right\} \pmod{m}.$$

These yield

$$a_p \equiv p a_1 \pmod{m}$$

and

$$(C) \qquad n a_1 \equiv 0 \pmod{m}.$$

The invariant operator is now representable in the form

$$(\xi_1 \xi_2^2 \cdots \xi_{n-1}^{n-1})^{a_1},$$

a_1 being a root of the congruence C. Every combination of the generators ξ_1 to ξ_{n-1} which is invariant in G must take this form. Conversely every such operator is invariant in G and the above representation is necessary and sufficient.

If we were to consider a combination involving all the ξ's but ξ_a, the reasoning would be identical with the above and it would be found that an invariant operator is necessarily of the form

$$(\xi_{a+1} \xi_{a+2}^2 \cdots \xi_n^{n-a} \xi_1^{n-a+1} \cdots \xi_{a-2}^{n-2} \xi_{a-1}^{n-1})^{a_1},$$

where again a_1 satisfies the congruence

$$n a_1 \equiv 0 \pmod{m}.$$

The above operator may be written

$$(\xi_{a+1} \xi_{a+2} \cdots \xi_{a-1})^{a_1} (\xi_{a+2} \xi_{a+3}^2 \cdots \xi_n^{n-a-1} \xi_1^{n-a} \cdots \xi_{a-1}^{n-2})^{a_1},$$

and by virtue of the relation

$$\prod_{i=1}^{n} \xi_i = 1$$

becomes

$$(\xi_{a+2} \xi_{a+3}^2 \cdots \xi_n^{n-a-1} \xi_1^{n-a} \cdots \xi_{a-1}^{n-2})^{a_1} \xi_a^{-a_1}.$$

But

$$-a_1 \equiv (n-1) a_1 \pmod{m},$$

so that

$$\begin{aligned}
&(\xi_{a+1} \xi_{a+2}^2 \cdots \xi_n^{n-a} \xi_1^{n-a+1} \cdots \xi_{a-2}^{n-2} \xi_{a-1}^{n-1})^{a_1} \\
&\quad = (\xi_{a+2} \xi_{a+3}^2 \cdots \xi_n^{n-a-1} \xi_1^{n-a} \cdots \xi_{a-1}^{n-2} \xi_a^{n-1})^{a_1} \\
&\quad = (\xi_1 \xi_2^2 \cdots \xi_{n-1}^{n-1})^{a_1}.
\end{aligned}$$

This shows that all the operators

$$(\xi_{a+1}\xi_{a+2}^{2}\cdots\xi_{n}^{n-a}\xi_{1}^{n-a+1}\cdots\xi_{a-1}^{n-1})^{a_1}\quad(a=0,1,\cdots,n-1)$$

are identical, and that one may with perfect generality consider

$$(\xi_1\xi_2^{2}\cdots\xi_{n-1}^{n-1})^{a_1}$$

as being the only permissible combination. As a result, the invariant operators of G contained in H form a cyclic sub-group whose order is determined as soon as the possible values of $a_1<m$ are known.

To find these values it is necessary to return to the congruence

$$na_1\equiv 0\pmod{m}.$$

If m is prime to n, $a_1=0$ is the only solution which is less than m. In such a case the identity is the only invariant operator. If m and n have the greatest common divisor d, the congruence reduces to

$$\frac{n}{d}a_1\equiv 0\quad\left(\mathrm{mod}\ \frac{m}{d}\right).$$

Here a_1 may take on all the values mq/d $(q=1,2,\cdots,d)$ and the sub-group is of order d. Both of these cases are combined in the general result that the number of invariant operators in H is d, where d is the greatest common divisor of m and n; they are all expressible in the form

$$(\xi_1\xi_2^{2}\cdots\xi_{n-1}^{n-1})^{mq/d}\quad(q=1,2,\cdots,d).$$

The generalization to any value of k follows along the same lines. Suppose some combination of ξ_1 to ξ_{n-k} to be invariant. For convenience, it is set down in the form

$$(\xi_1^{a_1}\xi_{k+1}^{a_2}\cdots\xi_{n-2k+1}^{a_{x-1}})(\xi_2^{b_1}\xi_{k+2}^{b_2}\cdots\xi_{n-2k+2}^{b_{x-1}})\cdots(\xi_k^{k_1}\xi_{2k}^{k_2}\cdots\xi_{n-k}^{k_{x-1}}),$$

where the operators of each set are kept together. If it is equal to its transform,

$$\begin{aligned}(\xi_1^{a_1}\xi_{k+1}^{a_2}\cdots\xi_{n-2k+1}^{a_{x-1}})&(\xi_2^{b_1}\xi_{k+2}^{b_2}\cdots\xi_{n-2k+2}^{b_{x-1}})\cdots(\xi_k^{k_1}\xi_{2k}^{k_2}\cdots\xi_{n-k}^{k_{x-1}})\\ &=(\xi_n^{a_1}\xi_k^{a_2}\cdots\xi_{n-2k}^{a_{x-1}})(\xi_1^{b_1}\xi_{k+1}^{b_2}\cdots\xi_{n-2k+1}^{b_{x-1}})\cdots(\xi_{k-1}^{k_1}\xi_{2k-1}^{k_2}\cdots\xi_{n-k-1}^{k_{x-1}}),\end{aligned}$$

from which

$$\begin{aligned}\xi_{n-k}^{k_{x-1}}=(\xi_n^{a_1}\xi_k^{a_2-k_1}\cdots\xi_{n-2k}^{a_{x-1}-k_{x-2}})&(\xi_1^{b_1-a_1}\xi_{k+1}^{b_2-a_2}\cdots\xi_{n-2k+1}^{b_{x-1}-a_{x-1}})\\ &\cdots(\xi_{k-1}^{k_1-j_1}\xi_{2k-1}^{k_2-j_2}\cdots\xi_{n-k-1}^{k_{x-1}-j_{x-1}}).\end{aligned}$$

Treating the first parenthesis as before and simplifying,

$$\xi_{n-k}^{a_1+k_{x-1}} = (\xi_k^{a_2-k_1-a_1}\xi_{2k}^{a_3-k_2-a_1}\cdots\xi_{n-2k}^{a_{x-1}-k_{x-2}-a_1})(\xi_1^{b_1-a_1}\xi_{k+1}^{b_2-a_2}$$
$$\cdots\xi_{n-2k+1}^{b_{x-1}-a_{x-1}})\cdots(\xi_{k-1}^{k_1-j_1}\xi_{2k-1}^{k_2-j_2}\cdots\xi_{n-k-1}^{k_{x-1}-j_{x-1}}).$$

The above relation involves only $n-k$ of the ξ's and $x-1$ of each set. Hence these $n-k$ are all independent and both members reduce to the identity. As a result a series of congruences is obtained:

$$\begin{aligned}
a_1 &\equiv b_1 \equiv c_1 \equiv \cdots \equiv k_1,\\
a_2 &\equiv b_2 \equiv c_2 \equiv \cdots \equiv k_2,\\
&\cdots\cdots\cdots\cdots\\
a_{x-1} &\equiv b_{x-1} \equiv c_{x-1} \equiv \cdots \equiv k_{x-1},\\
k_{x-1} &\equiv -a_1,\\
a_2 &\equiv k_1 + a_1,\\
a_3 &\equiv k_2 + a_1,\\
&\cdots\cdots\\
a_{x-1} &\equiv k_{x-2} + a_1 \qquad \pmod{m}.
\end{aligned}$$

These lead to the relations

$$\begin{aligned}
a_2 &\equiv 2a_1,\\
a_3 &\equiv 3a_1,\\
&\cdots\cdots\\
a_{x-1} &\equiv (x-1)a_1 \equiv -a \qquad \pmod{m},
\end{aligned}$$

so that finally

$$xa_1 \equiv 0 \qquad \pmod{m}.$$

From this point on, the reasoning is exactly like the preceding except that the quantity n is now replaced by $x=n/k$. Hence the sub-group of the central which is contained in H is cyclic and of order D where D is the greatest common divisor of m and n/k. The final form of the invariant operators is

$$[(\xi_1\xi_{k+1}^2\cdots\xi_{n-2k+1}^{x-1})(\xi_2\xi_{k+2}^2\cdots\xi_{n-2k+2}^{x-1})$$
$$\cdots(\xi_k\xi_{2k}^2\cdots\xi_{n-k}^{x-1})]^{mq/D} \qquad (q=1,2,\cdots,D)$$

and these operators will henceforth be denoted by h_i.

The question now arises "Are there any invariant operators in G which are not contained in H?" It will first be shown that if such operators exist

they must be of the form $S_1^v h_j$ where S_1^v is invariant in G. For, every operator outside of H is of the form

$$S_1^v \mathcal{H},$$

where $\mathcal{H}$ is an operator in H. If such an operator is invariant in G it is invariant under S_1, which transforms it into

$$S_1^v(S_1^{-1}\mathcal{H}S_1),$$

and hence

$$\mathcal{H} = S_1^{-1}\mathcal{H}S_1.$$

Consequently

$$\mathcal{H} = h_j.$$

If now the transform of $S_1^v h_j$ under S_2 is considered it follows that

$$S_2^{-1}S_1^vS_2 = S_1^v.$$

Hence the only additional invariant operators which need be sought are powers of S_1. If no power of S_1 is invariant under S_2, the central will be wholly contained in H and will be identical with the cyclic sub-group already found.

The next step then is to investigate when a relation such as

$$S_2^{-1}S_1^vS_2 = S_1^v$$

is satisfied. If it holds, so will

$$S_2^{n-v}S_1^vS_2^{-n+v} = S_1^v,$$

from which modified form it is possible to deduce a first necessary condition. For it implies

$$S_2^{n-v}S_1^v = (S_1^{n-v}S_2^v)^{-1} = S_1^vS_2^{n-v}.$$

In terms of the ξ's, this becomes

$$\prod_{i=1}^{n-v}\xi_i^{-1} = \prod_{i=1}^{v}\xi_i,$$

$$\prod_{i=n-v+1}^{n}\xi_i = \prod_{i=1}^{v}\xi_i,$$

and therefore

$$n - v = v,$$

whence

$$v = \frac{n}{2} \cdot$$

That is, the only power of S_1 which can be invariant in G is $S_1^{n/2}$.

If $S_1^{n/2}$ is invariant, then so is $S_1^{n/2}S_2^{n/2}$ also. For it is in H and is transformed into itself by S_2. But

$$S_1^{n/2}S_2^{n/2} = \prod_{i=1}^{n/2} \xi_i$$

and is invariant only when $x=2$; $m=D$. The latter of these relations requires m to be a divisor of n/k; the former requires that $k=n/2$.

These necessary conditions are also sufficient, and the resulting theorem is

THEOREM IV. *The central of G is cyclic and of order D where D is the greatest divisor of m and n/k except in the special case*

$$k = \frac{n}{2}; \quad m = 2;$$

in this latter case the order of the central is 2D.

The quotient group of H under G is cyclic and therefore abelian, hence H contains the commutator sub-group of G. To determine this sub-group, consider first the case $k=1$, and the following set of operators in H:

$$\xi_1\xi_2^{-1}, \ \xi_1\xi_3^{-1}, \cdots, \ \xi_1\xi_n^{-1}.$$

Each of these operators is a commutator; for

$$\xi_1\xi_a^{-1} = S_1S_2^{a-1}S_1^{-1}S_2^{-(a-1)}.$$

Again, the first $n-2$ of them are at once seen to be independent since they involve only $n-1$ of the ξ's. The group generated by them is therefore of order m^{n-2}. If the operator $\xi_1\xi_n^{-1}$ or any of its powers were in this group, they would be obtainable from the continued product of the first $n-2$ generators; for, it has already been seen that the relation between the ξ's involved all of them in a continued product. Now

$$(\xi_1\xi_2)^{-1}(\xi_1\xi_3^{-1}) \cdots (\xi_1\overset{-1}{\xi_{n-1}}) = \xi_1^{n-2}(\xi_2^{-1}\xi_3^{-1} \cdots \overset{-1}{\xi_{n-1}}) = \xi_1^{n-1}\xi_n.$$

Suppose

$$(\xi_1\xi_n^{-1})^{\alpha} = (\xi_1^{n-1}\xi_n)^{\beta}.$$

Then

$$\xi_1^{\alpha}\xi_n^{-\alpha} = \xi_1^{n\beta-\beta}\xi_n^{\beta},$$

from which

$$\alpha \equiv -\beta \pmod{m}$$

and

$$n\beta \equiv 0 \pmod{m}.$$

This last result is identical with the congruence obtained in the study of the central. The smallest value of β which satisfies it is m/d where d is the greatest common divisor of m and n. Hence $\xi_1\xi_n^{-1}$ is contained in the sub-group under consideration if and only if $d=m$. Omitting that case for the time being, it follows that the $n-1$ operators $\xi_1\xi_a{}^{-1}$ are independent and generate a group K of index d under H.

It will now be shown that this group K is the commutator sub-group, by demonstrating that every commutator is contained in it. However it is first necessary to obtain some preliminary results.

Suppose $f(\xi)$ represents any combination of the ξ_i, which by virtue of commutativity may be written in the form

$$\xi_1{}^{a_1}\xi_2{}^{a_2}\cdots\xi_n{}^{a_n}.$$

Denote by $f_{(-q)}(\xi)$ the combination

$$\xi_{1-q}^{a_1}\xi_{2-q}^{a_2}\cdots\xi_{n-q}^{a_n} \qquad (a-q\equiv n-a+q).$$

Then

$$f(\xi)\cdot[f_{(-q)}(\xi)]^{-1}$$

is in K. For this product contains n pairs of factors,

$$\xi_r^{a_r}\xi_{r-q}^{-a_r}$$

which may be written

$$(\xi_1\xi_r^{-1})^{-a_r}(\xi_1\xi_{r-q}^{-1})^{a_r}.$$

Consider now a general commutator of G. Every operator of G is expressible as some operator in H multiplied by a power of S_1 so that the most general commutator is

$$H_iS_1^{\alpha}\cdot H_jS_1^{\beta}(H_iS_1^{\alpha})^{-1}(H_jS_1^{\beta})^{-1} = H_iS_1^{\alpha}H_jS_1^{-\alpha}S_1^{\beta}H_i^{-1}S_1^{-\beta}H_j^{-1}.$$

It has already been seen that

$$S_1^{-1}\xi_iS_1 = \xi_{i-1},$$

from which

$$S_1^{-\alpha}\xi_i S_1^{\alpha} = \xi_{i-\alpha};$$

also

$$S_1\xi_i^{-1}S_1^{-1} = \xi_{i-1}^{-1},$$

$$S_1^{\alpha}\xi_i^{-1}S_1^{-\alpha} = \xi_{i-\alpha}^{-1}.$$

The above results show that the transform of any combination of ξ's by S_1^{α} reduces every original subscript by α. That is, it changes $f(\xi)$ into $f_{(-\alpha)}(\xi)$. The quantity

$$H_i(S_1^{\alpha}H_jS_1^{-\alpha})(S_1^{\beta}H_i^{-1}S_1^{-\beta})H_j^{-1}$$

is therefore equal to

$$H_iH_{j(-\alpha)}H_{i(-\beta)}^{-1}H_j^{-1} = H_iH_{i(-\beta)}^{-1}H_{j(-\alpha)}H_j^{-1}.$$

According to the preliminary lemma, this is contained in K and therefore K is the commutator sub-group.

Returning now to the case $d=m$, a group of index d under H would be generated by $n-2$ operators, thus explaining why $\xi_1\xi_n^{-1}$ is in this case contained in the group generated by the other $n-2$ of the operators $\xi_1\xi_a^{-1}$.

This remark now makes the result perfectly general for the case $k=1$. The commutator sub-group for every group of order nm^{n-1} is of index d under H. In particular, when $d=1$, i.e., when G contains no other invariant operator than the identity, the commutator sub-group coincides with H.

The generalization to any value of k offers no essential difficulty. The reasoning follows along the same lines and in the final result the only change is the replacement of d by D where D is the greatest common divisor of m and n/k.

THEOREM V. *The commutator sub-group of G is contained in H and is of index D where D is the greatest common divisor of m and n/k. It is generated by the $n-1$ operators*

$$\xi_1\xi_a^{-1} \qquad (a = 2, 3, \cdots, n).$$

VI. Special cases

The following special cases seem worthy of mention.

In the simplest case, $m=1$ and G is cyclic. $S_1=S_2$ and the relation

$$(S_1S_2^{k-1})^x = 1$$

reduces to an identity.

The family of groups G includes the dihedral groups as the special case $n=2$.

In the special case $n=3$, $k=1$, G is the group of order $3m^2$, previously studied by Edington* in his thesis and also by Miller.*

In the special case $n=4$, $k=2$, G is of order $4m^2$ and is another of the families obtained by Edington in his thesis. He defined the group by means of the relations

$$S_1^4 = S_2^4 = (S_1S_2)^2 = 1.$$

The condition $(S_1\,S_2)^2=1$ is exactly what $(S_1S_2^{k-1})^x=1$ reduces to on setting $x=2$. It is interesting to note in connection with this family of groups that it is not necessary to assume the operators $S_1^{\alpha}S_2^{n-\alpha}$ commutative. In this particular case that property follows as a consequence of the defining relations.

Finally, Edington's groups of order nm^{n-1} mentioned in the introduction are simply isomorphic with the groups obtained on setting $k=1$.

* W. E. Edington, these Transactions, vol. 25 (1923), p. 193. G. A. Miller, Proceedings of the National Academy of Sciences, vol. 13 (1927), p. 24.

George Washington University,
Washington, D. C.

GROUPS $\{S, T\}$ WHOSE COMMUTATOR SUBGROUPS ARE ABELIAN*

BY

H. R. BRAHANA

The groups generated by S and T satisfying the relations $S^3 = T^2 = (ST)^6 = 1$ were classified by Professor Miller.† The fact that makes these groups particularly easy to manage is that the commutator subgroups are abelian. It has been noted‡ that, with the exception of the tetrahedral group and the dihedral group of order 6, the only groups generated by an operator S of order 3 and an operator T of order 2 whose commutator subgroups are abelian are those considered by Miller. The groups which we shall consider are generated by S and T which satisfy the relations

$$S^p = T^2 = 1$$

where p is a prime, and have abelian commutator subgroups. The groups generated by two operators of order two are the well known dihedral groups. In view of this fact and of Miller's paper we may assume that p is a prime greater than 3.

1. Using the notation we have used before§ we let $\sigma_i = TS^{-i}TS^i$, $i = 1, 2, \cdots, p-1$. We note that T transforms σ_i into its inverse and that S transforms σ_i into $\sigma_1^{-1}\sigma_{i+1}$. The group generated by the σ's is therefore invariant; it is contained in the commutator subgroup H, and we shall show that it coincides with H.

Any operator of $\{S, T\}$ may be written in one of the forms

$$\sigma, \tag{1.1}$$

$$\sigma T, \tag{1.2}$$

$$\sigma \cdot TS^i, \tag{1.3}$$

$$\sigma \cdot S^i, \tag{1.4}$$

where σ is an operator of the group $\{\sigma_1, \sigma_2, \cdots, \sigma_{p-1}\}$. For, multiplication on the right by S puts (1.1) and (1.2) into (1.4) and (1.3) respectively, and either leaves the latter two in their present forms or puts one or both of them into

* Presented to the Society, August 31, 1932; received by the editors July 28, 1932.

† Quarterly Journal, vol. 33 (1901), pp. 76–79.

‡ American Journal of Mathematics, vol. 50 (1928), p. 347. The two groups of orders 6 and 12 should be excepted in that paper too.

§ Cf. the last reference.

(1.2) and (1.1) respectively. Multiplication on the right by T interchanges the first two and interchanges the last two, for $\sigma\cdot TS^iT=\sigma TS^iTS^{-i}\cdot S^i=\sigma\sigma_i\cdot S^i$ and $\sigma S^iT=\sigma S^iTS^{-i}T\cdot TS^i=\sigma\sigma_{-i}^{-1}\cdot TS^i$.

Now let $Q=\sigma T^{a_0}S^{a_1}$ and $R=\sigma'\cdot T^{b_0}S^{b_1}$, where σ and σ' are two operators of $\{\sigma_1, \sigma_2, \cdots, \sigma_{p-1}\}$, a_0 and b_0 are either or both 0 or 1, and a_1 and b_1 are integers less than p, be any two operators of $\{S, T\}$. Their commutator $Q^{-1}R^{-1}QR$ may be reduced as follows:

$$\begin{aligned}
\Sigma &= Q^{-1}R^{-1}QR \\
&= S^{-a_1}T^{a_0}\sigma^{-1}\cdot S^{-b_1}T^{b_0}\sigma'^{-1}\cdot\sigma T^{a_0}S^{a_1}\cdot\sigma' T^{b_0}S^{b_1} \\
&= S^{-a_1}T^{a_0}S^{-b_1}\sigma'' T^{b_0}\sigma'^{-1}\cdot\sigma T^{a_0}S^{a_1}\cdot\sigma' T^{b_0}S^{b_1}, \text{ where } \sigma'' = S^{b_1}\sigma^{-1}S^{-b_1}, \\
&= S^{-a_1}T^{a_0}S^{-b_1}T^{b_0}\sigma''' T^{a_0}S^{a_1}\cdot\sigma' T^{b_0}S^{b_1}, \text{ where } \sigma''' = \sigma''^{-1}\sigma'^{-1}\sigma, \\
&= S^{-a_1}T^{a_0}S^{-b_1}T^{b_0}\cdot T^{a_0}S^{a_1}\sigma^{\mathrm{iv}}T^{b_0}S^{b_1}, \text{ where } \sigma^{\mathrm{iv}} = S^{-a_1}T^{a_0}\sigma''' T^{a_0}S^{a_1}\cdot\sigma';
\end{aligned}$$

$$(1.5)\quad \Sigma = S^{-a_1}T^{a_0}S^{-b_1}T^{b_0}\cdot T^{a_0}S^{a_1}T^{b_0}S^{b_1}\cdot\sigma^{\mathrm{v}}, \text{ where } \sigma^{\mathrm{v}} = S^{-b_1}T^{b_0}\sigma^{\mathrm{iv}}T^{b_0}S^{b_1}.$$

Now if a_0 and b_0 are both zero, the right side of (1.5) is σ^{v}.

If $a_0=b_0=1$, (1.5) becomes $\sigma_{a_1}^{-1}\sigma_{b_1}\sigma^{\mathrm{v}}$.
If $a_0=0$, $b_0=1$, it becomes $\sigma_{a_1+b_1}^{-1}\cdot\sigma_{b_1}\sigma^{\mathrm{v}}$.
If $a_0=1$, $b_0=0$, it becomes $\sigma_{a_1}^{-1}\sigma_{a_1+b_1}\cdot\sigma^{\mathrm{v}}$.

So in any case Σ is an operator of $\{\sigma_1, \sigma_2, \cdots, \sigma_{p-1}\}$.

In the above we have not assumed that H was abelian nor that the order of S was prime. Hence,

(1.6) *In any group $\{S, T\}$ the commutator subgroup is generated by commutators of T and powers of S.*

2. We now assume that H is abelian and that the order of S is a prime number p. Since $T\sigma_iT=\sigma_i^{-1}$ it follows that T is in H only if H is of order 2^m and type $1, 1, \cdots$. Conversely, if H is abelian, of order 2^m, and type $1, 1, \cdots$, T is permutable with every operator of H, in which case T may or may not be in H as we shall see in a later section. We shall suppose hereafter, except where the contrary is explicitly stated, that T is not in H.

If S is in H the group $\{S, T\}$ is generalized dihedral, being generated by an abelian group H and an operator T which transforms every operator of H into its inverse. In fact $\{S, T\}$ must be the dihedral group of order $2p$. These two types of group are special and admit of special treatment. We shall then assume that neither S nor T is in H.

The quotient group of $\{S, T\}$ with respect to H is necessarily abelian. It is generated by two operators of orders p and 2 corresponding to S and T and therefore it must be cyclic and of order $2p$. The operator $(ST)^{2p}$ written

in terms of commutators* is

$$(ST)^{2p} = \sigma_{p-1}^{-1}\sigma_{p-2}\sigma_{p-3}^{-1}\sigma_{p-4} \cdots \sigma_1\sigma_{p-1}\sigma_{p-2}^{-1}\sigma_{p-3} \cdots \sigma_1^{-1}.$$

Since H is abelian this is identity. Hence,

(2.1) *If the abelian commutator subgroup contains neither S nor T, then (ST) is of order $2p$.*

3. We proceed to an examination of the $p-1$ σ's which generate H. We note first a relation which has been used before:

$$\begin{aligned} S^{-1}\sigma_i S &= S^{-1}\cdot TS^{-i}TS^i\cdot S = S^{-1}TST\cdot TS^{-(i+1)}TS^{i+1} \\ &= \sigma_1^{-1}\sigma_{i+1}. \end{aligned} \tag{3.1}$$

In the same way we may obtain

$$S^{-k}\sigma_i S^k = \sigma_k^{-1}\cdot\sigma_{i+k}, \tag{3.2}$$

where i and k may take on all the values $0, 1, 2, \cdots, p-1$, and $i+k$ is reduced modulo p.

If in (3.2) we take i to be 1 and allow k to take on the values $1, 2, \cdots, p-1$ we get the set of p conjugates of σ_1 under S. This set generates a group which contains σ_i for every i. Hence we have

(3.3) *The group H is generated by the set of conjugates of σ_1 under S.*

The $p-1$ σ's generate H, (1.6), and consequently H can have no more than $p-1$ independent generators. There must then be a relation connecting the p conjugates of σ_1 under S. There may be a relation connecting a smaller number of these conjugates. If there is a relation connecting the first $m+1$ conjugates under S, viz. $\sigma_1, \sigma_1^{-1}\sigma_2, \sigma_2^{-1}\sigma_3, \cdots, \sigma_m^{-1}\sigma_{m+1}$, then σ_{m+1} is in the group generated by the first m σ's. By (3.1) σ_{m+2} can be expressed in terms of the preceding σ's and hence in terms of the first m σ's. Therefore,

(3.4) *If $m+1$ is the smallest number of successive conjugates of σ_1 under S so that the last one may be expressed in terms of the preceding ones, then H has m independent generators.*

The p conjugates of σ_1 which by (3.3) generate H are of course of the same order, the order of σ_1; we shall show also that the $p-1$ σ's are of the same order.

Let the order of σ_i be n_i. If in (3.2) we let $i=p-k$ we have

$$S^{-k}\sigma_{p-k}S^k = \sigma_k^{-1}\cdot\sigma_p = \sigma_k^{-1}. \tag{3.5}$$

This implies that $n_k=n_{p-k}$. If in (3.1) we let $i=1$, we have $S^{-1}\sigma_1^n S$

* Cf. the reference given in the third footnote in this paper.

$=\sigma_1^{-n}\sigma_2^{n}$. When $n=n_1$ this gives $\sigma_2^{n_1}=1$, and therefore n_2 is a divisor of n_1. Similarly, allowing i in (3.1) to take on successively the values 2, 3, $\cdots$ it follows that n_i is a divisor of n_1.

If in (3.2) we let $k=2$, and let i take on successively the values 2, 4, 6, $\cdots$, $p-1$, we see in the same manner as above that $n_4, n_6, \cdots, n_{p-1}$ are divisors of n_2. By (3.5) we see that $n_{p-1}=n_1$, and since n_2 divides n_1 and $n_{p-1}=n_1$ divides n_2, it follows that $n_1=n_2=n_{p-1}=n_{p-2}$.

If in (3.2) we let $k=3$, and allow i to take on the values 3, 6, $\cdots$, we find that $n_6, n_9, \cdots$ are divisors of n_3. One of the numbers $p-1$ and $p-2$ is divisible by 3 and therefore $n_1=n_{p-1}=n_{p-2}$ divides n_3, and $n_3=n_1$.

This induction may be completed by showing in the same manner that if $n_i=n_1$, for $i=1, 2, \cdots, k$, then $n_{k+1}=n_1$. Hence

(3.6) *If the order of S is a prime and H is abelian, the σ's are all of the same order.*

4. In order to investigate the group H it is convenient to consider the orders of the operators in the co-set HS. The order of $\sigma_i S$ may be obtained as follows:

$$(4.1)\qquad \begin{aligned}(\sigma_i S)^p &= \sigma_i S\cdot\sigma_i S\cdot\sigma_i S\cdots\sigma_i S\cdot\sigma_i S\\ &= S^p\cdot S^{-p}\sigma_i S^p\cdot S^{-(p-1)}\sigma_i S^{p-1}\cdots S^{-2}\sigma_i S^2\cdot S^{-1}\sigma_i S.\end{aligned}$$

By application of (3.2) this becomes

$$(\sigma_i S)^p = S^p\cdot\sigma_i\cdot\sigma_{p-1}^{-1}\sigma_{i+p-1}\cdot\ \cdots\ \cdot\sigma_2^{-1}\sigma_{i+2}\cdot\sigma_1^{-1}\sigma_{i+1}.$$

The σ's with negative exponents are $\sigma_1, \sigma_2, \cdots, \sigma_{p-1}$. Those with positive exponents are p in number including $\sigma_{i+p-i}=\sigma_p=1$. The remaining $p-1$ are $\sigma_1, \sigma_2, \cdots, \sigma_{p-1}$. Hence we have

$$(4.2)\qquad (\sigma_i S)^p = 1.$$

If we take the pth power of any operator of HS we have

$$(4.3)\qquad (\sigma_1^{\alpha_1}\sigma_2^{\alpha_2}\cdots\sigma_{p-1}^{\alpha_{p-1}}S)^p = (\sigma_2^{\alpha_2}\sigma_3^{\alpha_3}\cdots\sigma_{p-1}^{\alpha_{p-1}}S)^p.$$

This is obtained by moving $\sigma_1^{\alpha_1}$ past successive S's by means of (3.1) in the form $\sigma_i S=S\cdot\sigma_1^{-1}\sigma_{i+1}$. By repeated application of (4.3) we reduce the pth power of any element of HS to (4.2) and hence obtain

(4.4) *Every operator in the co-set HS is of order p.*

From this theorem we may draw many important conclusions concerning H and S. We note first

(4.5) *Any operator of H which is permutable with S is of order p.*

From (4.5) and the fact that T transforms every operator of H into its inverse follows

(4.6) *The central of* $\{S, T\}$ *is identity.*

If H contains operators of order p then $\{\sigma_1\}$ contains operators of order p. If an operator of order p in $\{\sigma_1\}$ is not invariant under S the cyclic group generated by it is not invariant and contains no invariant operator except identity. Its conjugates generate a group of order p^m, $m \leqq p-1$, and type 1, 1, $\cdots$. This group contains $1+p+p^2+\cdots+p^{m-1}$ subgroups and at least one of them is invariant under S.

5. H is the direct product of its Sylow subgroups, and each of its Sylow subgroups is invariant under S. Let the order of H be $p^a p_1^{a_1} p_2^{a_2} p_3^{a_3} \cdots$, where the p's are distinct primes. Then by (4.5) none of the Sylow subgroups corresponding to p_i, $i=1, 2, 3, \cdots$, can contain invariant operators. Hence

(5.1) *If the order of* H *is* $p^a p_1^{a_1} p_2^{a_2} \cdots$, *where the* p'*s are distinct primes, each of the numbers* $p_i^{a_i}$ *is congruent to* 1, mod p.

If the residue of p_i, mod p, belongs to the exponent $p-1$, then the number a_i must be at least $p-1$. The Sylow subgroup H_{p_i} of order p_i contains a characteristic subgroup composed of operators of order p_i contained in subgroups generated by operators of highest order in H_{p_i}. When p_i belongs to the exponent $p-1$ the order of this characteristic subgroup must be p_i^{p-1}, and therefore H_{p_i} must contain $p-1$ independent generators of highest order.

If α_i is used to denote the exponent to which p_i belongs, mod p, then α_i is a divisor of $p-1$. Replacing the exponent $p-1$ in the preceding paragraph by α_i we have the result that the number of independent generators of highest order of H_{p_i} is a multiple of α_i. Continuing the argument we observe that H_{p_i} contains a characteristic subgroup generated by operators of order p_i contained in cyclic subgroups of next to highest order in H_{p_i}. The order of this subgroup must also be a multiple of $p_i^{\alpha_i}$, and the number of independent generators of H_{p_i} of next to highest order must be a multiple of α_i. This argument may be continued to give

(5.2) *The number of independent generators of* H_{p_i} *of each order for which there is one is a multiple of* α_i, *the exponent to which* p_i *belongs modulo* p.

6. The Sylow subgroup H_{p_i} of H is invariant under H, S, and T, and is therefore an invariant subgroup of $\{S, T\}$. The corresponding quotient group of $\{S, T\}$ is generated by two operators of orders p and 2, and its commutator subgroup is abelian, being the quotient group of H with respect to H_{p_i}. Moreover, any product of the H_{p_i}'s is invariant in $\{S, T\}$ and the corresponding quotient group is of the same type as $\{S, T\}$. Hence,

(6.1) *The existence of the group* $\{S, T\}$ *whose commutator subgroup is abelian and of order* $p^{a}p_1^{a_1}p_2^{a_2}\cdots$ *implies the existence of a group* $\{S', T'\}$ *whose commutator subgroup is abelian and of order* $p_i^{a_i}$ *for each* i.

From this theorem it follows that the question of the existence of a group $\{S, T\}$ with a given abelian group H as a commutator subgroup must be concerned with the existence of groups $\{S, T\}$ whose commutator subgroups are Sylow subgroups of H. We shall show that the existence of $\{S, T\}$ follows from the existence of each of these groups whose commutator subgroups are prime power groups, thus proving the converse of (6.1).

Let $\{S', T'\}$ and $\{S'', T''\}$ be two groups whose abelian commutator subgroups H' and H'' are of orders $p_1^{a_1}$ and $p_2^{a_2}$ respectively, where p_1 and p_2 are distinct primes, let S' and S'' be of the same prime order p, and let σ_1' and σ_1'' be commutators of the respective pairs of generators. Now let H be the direct product of H' and H'', and let T be an operator of order 2 which transforms every operator of H into its inverse. The group $\{H, T\}$ is generalized dihedral and its group of isomorphisms is abstractly the same as the holomorph of H.* The group of isomorphisms of H is the direct product of the groups of isomorphisms of H' and H'' and hence the holomorph of H contains operators of order p. We wish to show that there is one such operator S which with T will generate a group having H as a commutator subgroup. Let S be an operator which transforms H' and H'' as they are transformed by S' and S'' respectively, and let $S^{-1}TS = T\sigma_1'\sigma_1''$. Then

$$
\begin{aligned}
S^{-k}TS^{k} &= S^{-(k-1)}T\sigma_1'\sigma_1'\,S^{k-1} = S^{-(k-1)}TS^{k-1}\sigma_{k-1}^{\prime-1}\sigma_k'\,\sigma_{k-1}^{\prime\prime-1}\sigma_k'',\\
&= S^{-(k-1)}TS^{k-1}\Sigma_k, \text{ where } \Sigma_m = \sigma_{m-1}^{\prime-1}\sigma_m'\,\sigma_{m-1}^{\prime\prime-1}\sigma_m'',\\
&= S^{-(k-2)}TS^{k-2}\Sigma_{k-1}\Sigma_k,\\
&\quad \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\\
&= S^{-1}TS\Sigma_2\cdots\Sigma_{k-1}\Sigma_k,\\
&= T\sigma_1'\sigma_1''\Sigma_2\Sigma_3\cdots\Sigma_{k-1}\Sigma_k.
\end{aligned}
$$

If $k = p$, we have

$$S^{-p}TS^{p} = T\sigma_1'\sigma_1''\Sigma_2\Sigma_3\cdots\Sigma_{p-1}\Sigma_p.$$

Taking account of the definition of Σ_m and of (4.1) and (4.2), we have $S^{-p}TS^{p} = T$. Since S transforms $\{H, T\}$ according to an operator of order p we may take S to be of order p†. The group $\{S, T\}$ contains $\sigma_1'\sigma_1''$, and since the orders of σ_1' and σ_1'' are relatively prime, contains both σ_1' and σ_1''. The

* Miller, Blichfeldt, and Dickson, *Finite Groups*, p. 169.

† American Journal of Mathematics, vol. 52 (1930), p. 919.

group generated by the conjugates of $\sigma_1'\sigma_1''$ under S contains the conjugates of σ_1' and σ_1'' under S' and S'' respectively and so is H, which by (1.6) is the commutator subgroup of $\{S, T\}$. Hence we have

(6.2) *The existence of two groups $\{S', T'\}$ and $\{S'', T''\}$ whose abelian commutator subgroups are of orders relatively prime and with S' and S'' of the same prime order p, implies the existence of a group $\{S, T\}$ with S of order p and a commutator subgroup which is the direct product of the commutator subgroups of the two given groups.*

7. We consider next the subgroups $\{S', T'\}$ which have abelian commutator subgroups and are generated by an operator of order p and one of order 2. Let $S' = \sigma_{p-1}^{l} S$ and $T' = \sigma_1^k T$. Then

$$
\begin{aligned}
\sigma_1' &= T'S'^{-1}T'S' = \sigma_1^k T S^{-1} \sigma_{p-1}^{-l} \sigma_1^k T \sigma_{p-1}^{l} S \\
&= \sigma_1^k T S^{-1} T \sigma_1^{-k} \sigma_{p-1}^{2l} S \\
&= \sigma_1^k T S^{-1} T S \cdot \sigma_1^k \sigma_2^{-k} \sigma_1^{-2l} \\
&= \sigma_1^{2k-2l+1} \sigma_2^{-k}.
\end{aligned}
\tag{7.1}
$$

Let the order of σ_1 be m. When m is odd k may be chosen to be m and then l may be chosen so that $1-2l$ is any number less than or equal to m. Then (7.1) becomes $\sigma_1' = \sigma_1^{m'}$, where m' is any number; hence l may be chosen so that m' is any divisor of m. If m is even then for any choice of k the number $2k-2l+1$ is still odd and for proper choice of k and l will be any odd divisor of m; or k may be chosen as any even number and then l may be chosen so that $2k-2l+1$ is any odd number less than m. If in the latter case k is taken to be the highest power of 2 contained in m and l chosen so that $2k-2l+1$ is the largest odd divisor of m, then σ_1' will be of even order. Hence,

(7.2) *If the order of H is $p^a 2^{a_1} p_1^{a_2} \cdots$, then $\{S, T\}$ contains subgroups $\{S', T'\}$ whose commutator subgroups H' have the orders $2^{a_1} p_2^{k_2 a_2} p_3^{k_3 a_3} \cdots$, where the k's are zeros or ones independently.*

8. In the preceding pages we have determined various conditions on H which are necessary in order that H be the abelian commutator subgroup of $\{S, T\}$. The last conditions are conditions on the order of H. We wish to inquire what conditions on H may be sufficient to determine H so that a group $\{S, T\}$ necessarily exists containing H as a commutator subgroup.

We proceed to prove the following fundamental theorem:

(8.1) *Necessary and sufficient conditions that there exist a group $\{S, T\}$ generated by S of order p and T of order 2 and containing a given abelian group*

H as a commutator subgroup are (1) *that the group of isomorphisms of H contain an operator U of order p whose powers transform an operator s_1 of H into a set of generators, and* (2) *that the product of the operators in the set of conjugates under U which contains s_1 be identity.*

The conditions are necessary because of (3.3) and (4.2). Conversely, the group of isomorphisms of H contains an operator of order 2 which transforms every operator of H into its inverse; let T be an operator of order 2 which performs this transformation. Let S be an operator which transforms H according to U and which transforms T into Ts_1. The order of S has not yet been determined, but if its pth power transforms T into itself it will be possible to require further that S be of order p*. We therefore determine $S^{-p}TS^p$.

$$S^{-k}TS^k = S^{-(k-1)}Ts_1S^{k-1} = S^{-(k-1)}TS^{k-1}\cdot S^{-(k-1)}s_1S^{k-1}.$$

If we denote the successive conjugates of s_1 under U as follows:

$$U^{-k}s_1U^k = s_{k+1},$$

we have

$$S^{-p}TS^p = Ts_1s_2 \cdots s_{p-1}s_p.$$

If $s_1s_2 \cdots s_{p-1}s_p = 1$, then S^p is permutable with T and S may be taken to be of order p, which completes the proof of the converse.

Since the two conditions above are sufficient to ensure the existence of $\{S, T\}$ it follows that the conditions stated in the preceding theorems hold; in particular (4.5) holds. The question arises as to whether or not (2) can be replaced by (4.5). The product of the set of conjugates of s_1 is of course invariant under U. If (4.5) holds and the order of H is q^n, where q is prime to p, this product must be identity. Hence,

(8.2) *If H is abelian and of order q^n, where q is prime to p, then condition* (1) *of* (8.1) *and* (4.5) *are necessary and sufficient that there exist a group $\{S, T\}$ having H as a commutator subgroup.*

If the order of H is p, it is obvious that condition (2) of (8.1) may not be replaced by (4.5), for the group of order p^p and type $1, 1, \cdots$ admits an automorphism U,

$$U^{-1}s_iU = s_{i+1},$$

such that s_1 and U satisfy the given conditions. But since H has p independent generators it cannot be the commutator subgroup of any group $\{S, T\}$. The groups H satisfying (1) of (8.1), (4.5), and having not more than $p-1$ inde-

* Cf. the second reference in §6.

pendent generators have been determined in another paper* and the groups in which we are interested will be found among them. They are of the two following categories:

(a) H of order $p^{k_1 m+k_2(m-1)}$ with k_1 independent generators of order p^m and k_2 of order p^{m-1}, where $k_1+k_2=p-1$ and $m\geqq 1$;

(b) H of order p^{k+2} and type 2, 1, 1, $\cdots$, where $k<p-2$.

In the case of the groups of the first category it is always possible to select U so that conditions (1) and (2) of (8.1) hold. Let us suppose that $k_1=p-1$, so that $k_2=0$ and H is of type $m, m, \cdots$. Let $s_1, s_2, \cdots, s_{p-1}$ be a set of independent generators of H, all necessarily of order p^m. Then the group of isomorphisms of H contains an operator U defined as follows:

$$\begin{aligned} U^{-1}s_iU &= s_{i+1}, \quad i = 1, 2, \cdots, p-2, \\ U^{-1}s_{p-1}U &= s_1^{-1}s_2^{-1}\cdots s_{p-1}^{-1}. \end{aligned} \tag{8.3}$$

It is obvious that the order of U is p, and the product of the set of conjugates of s_1 under U is identity.

The existence of a group $\{S, T\}$ for the H just considered follows from (8.1). H contains a subgroup H' of order p which is invariant in $\{S, T\}$. The quotient group of $\{S, T\}$ with respect to H' is, by an argument similar to that used to establish (6.1), of the kind we are considering and its commutator subgroup is the quotient group of H with respect to H'; it is of type $m, m, \cdots, m, m-1$. By taking successive quotient groups with respect to invariant subgroups of order p, we may obtain a group $\{\overline{S}, \overline{T}\}$ having any one of the groups of category (a) as a commutator subgroup.

The groups of category (b) do not admit of isomorphisms which with H satisfy conditions (1) and (2) of (8.1). Let H be of type 2, 1, 1, $\cdots$, and let s_1, an operator of H, be of order p^2. Then there exists an operator U of order p in the group of isomorphisms of H which transforms s_1 into a set of generators.† Now $U^{-1}s_1U=s_1s_\alpha$, where s_α is some operator of H. Because of the type of H, s_1^p must be invariant under U and hence s_α is of order p and may be taken to be s_2, one of a set of independent generators. We may then suppose the generators of H to be chosen so that

$$\begin{aligned} U^{-1}s_1U &= s_1s_2, \\ U^{-1}s_iU &= s_{i+1}, \quad i = 2, 3, \cdots, k, \\ U^{-1}s_{k+1}U &= s_1^{a_p}s_2^{a_2}\cdots s_{k+1}^{a_{k+1}}. \end{aligned} \tag{8.4}$$

* *Prime-power abelian groups generated by a set of conjugates under a special automorphism*, to be published in the American Journal of Mathematics.

† Ibid., (5.64).

As was shown* in the paper referred to, the numbers $a_2, a_3, \cdots, a_{k+1}$ are completely determined by k. In considering successive conjugates of s_1 under U we need only consider the powers of s_1 in these conjugates, as will be apparent at the conclusion. Each of the first $k+1$ conjugates contains s_1 to the first power. Successive ones thereafter have s_1 to the powers

$$(8.5)\qquad \begin{aligned} &1+ap, \\ &1+\left[1-\binom{r-1}{1}\right]ap, \\ &1+\left[1-\binom{r-1}{1}+\binom{r-1}{2}\right]ap, \end{aligned}$$

and so on, where $r=p-1-k$ and the numbers in the brackets are the binomial coefficients. The last one in the series (8.5) is

$$1+\left[1-\binom{r-1}{1}+\binom{r-1}{2}-\cdots+(-1)^{n-2}\binom{r-1}{r-2}\right]ap.$$

The sum of these numbers is readily obtained. We note first that

$$1-\binom{r-1}{1}+\binom{r-1}{2}-\cdots+(-1)^i\binom{r-1}{i}=(-1)^i\binom{r-2}{i}.$$

Then the coefficient of ap in the sum will be

$$1-\binom{r-2}{1}+\binom{r-2}{2}-\cdots+(-1)^{r-2}\binom{r-2}{r-2},$$

which is $(1-1)^{n-2}=0$. Hence the sum is simply the sum of the p numbers independent of ap, all of which are 1's. Therefore the product of the p conjugates in the set which contains s_1, contains $s_1{}^p$, and cannot be identity† since the s_i's are independent generators. Since any operator U of order p which transforms an operator s_1 of H into a set of generators can be written in the form (8.4), and since the product of the set of conjugates of s_1 is not identity, it follows that no group H of the second category is the commutator subgroup of a group $\{S, T\}$. We may then state the following theorem:

(8.6) *In order that there exist a group* $\{S, T\}$ *having a given abelian group* H *of order* p^n *as a commutator subgroup it is necessary and sufficient that* (1) $n=k_1m+k_2(m-1)$, (2) $k_1+k_2=p-1$, *and* (3) H *have* k_1 *and* k_2 *independent generators of orders* p^m *and* p^{m-1} *respectively.*

---

* Ibid., (5.26).

† The product is exactly $s_1{}^p$, but it is not necessary to prove it here.

The corresponding theorem for the case where the order of H is q^n, q prime to p, is obtained from (8.2) and (3.2) of the paper referred to above.

(8.7) *In order that there exist a group* $\{S, T\}$ *having a given abelian group* H *of order* q^n *as a commutator subgroup, it is necessary and sufficient that* (1) $n = \alpha(k_1m_1+k_2m_2+ \cdots +k_jm_j)$, *where* α *is the exponent to which* q *belongs,* mod p, (2) $k_1+k_2+ \cdots +k_j \leqq (p-1)/\alpha$, *and* (3) H *have* $k_i\alpha$ *independent generators of order* q^{m_i}.

On the basis of a knowledge of the elementary part of the theory of abelian groups, theorems (6.2), (8.6), and (8.7) give a complete solution of the problem of the determination of the groups designated in the introduction.

University of Illinois,
Urbana, Ill.

ON THE REPRESENTATION OF A POLYNOMIAL IN A GALOIS FIELD AS THE SUM OF AN EVEN NUMBER OF SQUARES*

BY

LEONARD CARLITZ†

1. **Introduction.** Let $GF(p^n)$ denote a fixed Galois field of order p^n, p being any *odd* prime, and n an arbitrary positive integer; let $\mathfrak{D}(x, p^n)$ denote the totality of polynomials in an indeterminate, x, with coefficients in $GF(p^n)$. In this paper we seek simple expressions for the number of representations of a polynomial in $\mathfrak{D}$ as a sum of squares of polynomials in $\mathfrak{D}$ that satisfy certain restrictions.

More precisely, suppose that F is a *primary* polynomial, that is, the coefficient of the highest power of x occurring in F is the 1 element of the Galois field. Let s be a positive integer; $\alpha_1, \cdots, \alpha_s, \beta_1, \cdots, \beta_s$, $2s$ elements of the Galois field such that

$$\gamma_i = \alpha_i + \beta_i \neq 0 \qquad (i = 1, \cdots, s).$$

Then

(A) If F is of even degree, $2k$, and

$$\gamma = \gamma_1 + \gamma_2 + \cdots + \gamma_s \neq 0,$$

we seek the number of solutions of

$$\gamma F = \alpha_1 X_1^2 + \beta_1 Y_1^2 + \cdots + \alpha_s X_s^2 + \beta_s Y_s^2 \tag{1}$$

in primary polynomials X_i, Y_i, each of degree k.

(B) If F is of *arbitrary* degree f, $2k$ is any even integer $>f$, α any non-zero element of the Galois field, and

$$\gamma_1 + \gamma_2 + \cdots + \gamma_s = 0,$$

we seek the number of solutions of

$$\alpha F = \alpha_1 X_1^2 + \beta_1 Y_1^2 + \cdots + \alpha_s X_s^2 + \beta_s Y_s^2 \tag{2}$$

in primary polynomials X_i, Y_i, each of degree k.

The solution of (A) is expressed in terms of one of the functions $\rho_t(F)$, $\omega_t(F)$, defined thus:

* Presented to the Society, August 30, 1932; received by the editors June 23, 1932.

† This paper was written when the author was an International Research Fellow at Cambridge University.

$$(3)\qquad \rho_t(F) = \left(1 - \frac{1}{p^{nt}}\right) \sum_{M|F}^{m>k} |M|^t + \sum_{M|F}^{m=k} |M|^t, \quad |M| = p^{nm},$$

where m denotes the degree of M: the first summation is over all primary M dividing F, and of degree $>k$, the second is over all primary M dividing F and of degree $=k$;

$$(4)\qquad \omega_t(F) = \left(1 + \frac{1}{p^{nt}}\right) \sum_{M|F}^{m>k} (-1)^m |M|^t + \sum_{M|F}^{m=k} (-1)^m |M|^t,$$

the summations having the same meaning as in (3). If now

$$(5)\qquad (-1)^s \alpha_1 \cdots \alpha_s \beta_1 \cdots \beta_s$$

is a square in $GF(p^n)$, then the number of solutions of (1) is $\rho_{s-1}(F)$; if the expression (5) is a non-square of $GF(p^n)$, then the number of solutions is $\omega_{s-1}(F)$.

Case (B) involves a modification of the ρ and ω functions: if (5) is a square, the number of solutions of (2) is

$$p^{n(s-1)(2k-f)} \rho_{s-1}^k(F),$$

where

$$(6)\qquad \rho_t^k(F) = \left(1 - \frac{1}{p^{nt}}\right) \sum_{M|F}^{m>f-k} |M|^t + \sum_{M|F}^{m=f-k} |M|^t;$$

if (5) is a non-square, the number of solutions is

$$p^{n(s-1)(2k-f)} \omega_{s-1}^k(F),$$

where

$$(7)\qquad (-1)^f \omega_t^k(F) = \left(1 + \frac{1}{p^{nt}}\right) \sum_{M|F}^{m>f-k} (-1)^m |M|^t + \sum_{M|F}^{m=f-k} (-1)^m |M|^t.$$

If $k>f$, the second sum in (6) and (7) is vacuous and denotes zero.

We first treat case (A); then, making use of the results for this case, it is easy to deduce the results (6) and (7) for case (B). The method used is quite elementary, and presupposes only some well known general theorems concerning Galois fields.*

It should be emphasized that the results of this paper hold for all positive s. This is rather surprising when comparison is made with the known results concerning the number of representations of an ordinary integer as the sum of $2s$ squares: in the latter problem, while the cases $2s=2, 4, 6, 8$ admit of

* These theorems will be found in Dickson's *Linear Groups*, 1901, pp. 3–54.

simple expressions in terms of divisor functions, this is no longer true for $2s>8$. While comparison of the problem of this paper with the ordinary problem is of some interest, actually, since we are considering representations in terms of *primary* polynomials, the analogy is closer with the question of the number of representations of an integer as the sum of squares of *positive* integers.

Finally, we remark that it is possible, by methods similar to those used here, to determine the number of representations of a polynomial by means of any *odd* number of squares (that satisfy certain conditions). As we shall show in another paper, the final formulas are of quite a different type; they are no longer functions of divisors but involve sums of quadratic characters.

2. **Notation; preliminary lemmas.*** We shall employ the following notation. Polynomials will be denoted by large italic letters; unless the contrary be explicitly stated, a polynomial will always be assumed primary. Ordinary integers will be denoted by small italic letters, elements of the Galois field by small Greek letters. The degree of a polynomial will be denoted by the corresponding small letter, and we shall write

$$f = \deg F, \quad |F| = p^{nf}.$$

(A, B) is the "greatest" common divisor of A and B.

Using this notation, we have the following lemmas.

LEMMA 1. *The number of sets of polynomials* $[A, B]$ *such that*

$$\deg A = a, \ \deg B = b, \ (A, B) = 1,$$

is

$$\psi(a, b) = \begin{cases} p^{a+b} - p^{a+b-1} & \textit{for } ab \neq 0, \\ p^{a+b} & \textit{for } ab = 0. \end{cases}$$

This lemma is a special case of a more general theorem to be proved elsewhere. For completeness we give the following simple proof. Let us classify the $p^{n(a+b)}$ sets of polynomials $[A, B]$, $\deg A = a$, $\deg B = b$, according to their g.c.d. Then, if $1 \leqq a \leqq b$,

$$p^{n(a+b)} = \sum_{m \leqq a} |M| \psi(a-m, b-m), \tag{8}$$

the sum being extended over all M of degree $\leqq a$. The right member of (8) is

$$\begin{aligned} \psi(a, b) + \sum_{m=1}^{a} p^{nm} \psi(a-m, b-m) \\ = \psi(a, b) + p^{n(a+b-1)}, \end{aligned}$$

whence the lemma.

* The results of §§2, 3 hold for all p.

LEMMA 2. *Let F, A, B be of degree f, a, b, respectively*; $(A, B)=1$.

(I) *If $a+b\leqq f$, and α and β are two non-zero elements of $GF(p^n)$ such that $\alpha+\beta\neq 0$, then the number of solutions of*

$$(\alpha+\beta)F = \alpha AU + \beta BV, \quad |AU| = |BV|, \tag{9}$$

in polynomials U, V, is $|F/(AB)| = p^{n(f-a-b)}$.

(II) *If k is an integer $>f$, $a+b\leqq k$, and α is any non-zero element of $GF(p^n)$, then the number of solutions of*

$$\alpha F = AU - BV, \quad |AU| = |BV| = p^{nk}, \tag{10}$$

is

$$p^{n(k-a-b)} = p^{n(k-f)}\cdot |F|\cdot|AB|^{-1}.$$

It will suffice to prove (II) alone. From (10), we have

$$\begin{aligned} U &\equiv A' \pmod{B}, \quad a' < b, \\ V &\equiv B' \pmod{A}, \quad b' < a, \end{aligned} \tag{11}$$

where A' and B' are not necessarily primary. Since

$$u = \deg U = k - a \geqq b, \text{ and } v \geqq a,$$

the congruences (11) may be written in the form

$$U = A' + BU', \; V = B' + AV' \qquad (u' = v' = k - a - b),$$

where now U' and V' are primary. Then (10) becomes

$$\frac{\alpha F - AA' - BB'}{AB} = U' - V'. \tag{12}$$

But since $(A, B)=1$, there is a unique pair of polynomials A', B', such that $a'<b$, $b'<a$, and the left member of (12) is integral. If then V' be any (primary) polynomial of degree $v'=k-a-b$, U' is uniquely determined, and, retracing the steps that led from (10) to (12), U, V are uniquely determined. Since V' can be chosen in

$$p^{nv'} = p^{n(k-a-b)}$$

ways, this proves case (II). The proof of case (I) is very much the same.

3. **Theorems on the ρ and ω functions.** We now prove certain formulas concerning the functions $\rho_t(F)$ and $\omega_t(F)$ defined in the Introduction. As we shall see in the next section, these formulas enable us to solve our problem concerning (1); furthermore, the formulas seem to be of some interest in themselves.

THEOREM 1. *If F is of even degree, $2k$; α, β two elements of $GF(p^n)$ such that $\alpha\beta(\alpha+\beta)\neq 0$; s, t two (real or complex) numbers; then*

$$\sum \rho_s(A)\rho_t(B) = \rho_{s+t+1}(F), \tag{13}$$

$$\sum \rho_s(A)\omega_t(B) = \omega_{s+t+1}(F), \tag{14}$$

$$\sum \omega_s(A)\omega_t(B) = \rho_{s+t+1}(F), \tag{15}$$

where, in each instance, the summation is extended over all (primary) polynomials A, B of degree $2k$, such that

$$(\alpha + \beta)F = \alpha A + \beta B.$$

The three formulas (13), (14), (15) may be proved simultaneously if ρ and ω be expressed in terms of the function $\Lambda_s(F, \lambda)$ now to be defined. We define the "character" $\lambda(B)$ by

$$\lambda(B) = (-1)^b, \quad b = \deg B, \tag{16}$$

and the function $\Lambda_s(F, \lambda^i)$ by

$$\Lambda_s(F, \lambda^i) = \left(1 - \frac{(-1)^i}{p^{ns}}\right) \sum_{M|F}^{m>k} \lambda^i(M)\,|\,M\,|^s + \sum_{M|F}^{m=k} \lambda^i(M)\,|\,M\,|^s. \tag{17}$$

It is obvious from the definitions (3) and (4) that

$$\rho_s(F) = \Lambda_s(F, 1), \quad \omega_s(F) = \Lambda_s(F, \lambda), \tag{18}$$

and therefore the several parts of Theorem 1 reduce to

$$\sum_{(\alpha+\beta)F=\alpha A+\beta B} \Lambda_s(A, \lambda^i)\Lambda_t(B, \lambda^j) = \Lambda_{s+t+1}(F, \lambda^{i+j}), \tag{19}$$

i, j integers which may be taken $=0$ or 1. We proceed to establish (19).

The left hand member of (19) is by (17)

$$\begin{aligned}\Bigg\{&\left(1 - \frac{(-1)^i}{p^{ns}}\right)\left(1 - \frac{(-1)^j}{p^{nt}}\right) \sum_{a<u,\, b<v} + \left(1 - \frac{(-1)^i}{p^{ns}}\right) \sum_{a<u,\, b=v} \\ &+ \left(1 - \frac{(-1)^j}{p^{nt}}\right) \sum_{a=u,\, b<v} + \sum_{a=u,\, b=v} \Bigg\} \lambda^i(U)\,|\,U\,|^s\, \lambda^j(V)\,|\,V\,|^t,\end{aligned} \tag{20}$$

where each summation is taken over all (primary) A, B, U, V satisfying $(\alpha+\beta)F = \alpha AU + \beta BV$ as well as the conditions indicated under each $\sum$. Call the sums $\sum_1$, $\sum_2$, $\sum_3$, $\sum_4$, respectively; then, since $a<u$ is equivalent to $a<k$,

$$\sum\nolimits_1 = |F|^{s+t} \sum_{\substack{(\alpha+\beta)F=\alpha AU+\beta BV \\ a,b<k}} \lambda^i(A)\lambda^j(B)|A|^{-s}|B|^{-t}$$

$$(21) \qquad = |F|^{s+t} \sum_{M|F}^{m<k} \lambda^{i+j}(M)|M|^{-s-t}S_M,$$

where

$$\begin{aligned} S_M &= \sum_{\substack{(\alpha+\beta)FM^{-1}=\alpha AU+\beta BV \\ (A,B)=1;a,b<k-m}} \lambda^i(A)\lambda^j(B)|A|^{-s}|B|^{-t} \\ &= \sum_{\substack{(A,B)=1 \\ a,b<k-m}} \lambda^i(A)\lambda^j(B)|A|^{-s}|B|^{-t} \sum_{(\alpha+\beta)FM^{-1}=\alpha AU+\beta BV} 1 \\ &= |FM^{-1}| \sum_{\substack{(A,B)=1 \\ a,b<k-m}} \lambda^i(A)\lambda^j(B)|A|^{-s-1}|B|^{-t-1}, \end{aligned}$$

by case (I) of Lemma 2. By the definition of $\psi(a, b)$ in Lemma 1, the last expression is equal to

$$(22) \qquad |FM^{-1}| \sum_{a,b<k-m} (-1)^{ia+jb} p^{-n(s+1)a-n(t+1)b}\psi(a, b).$$

Applying Lemma 1, the sum becomes

$$\sum_{a=0}^{k-m-1}(-1)^{ia}p^{-nsa} + \sum_{b=1}^{k-m-1}(-1)^{jb}p^{-ntb} + \sum_{a,b=1}^{k-m-1}(-1)^{ia+jb}p^{-n(sa+tb)}(1-p^{-n})$$
$$= [k-m, i]_s[k-m, j]_t - p^{-n}[k-m, i]_s'[k-m, j]_t',$$

where, for brevity, we put

$$(23) \qquad \begin{aligned} [k, i]_s &= \frac{1-(-1)^{ik}p^{-nsk}}{1-(-1)^i p^{-ns}},* \\ [k, i]_s' &= \frac{(-1)^i p^{-ns} - (-1)^{ik}p^{-nsk}}{1-(-1)^i p^{-ns}} = [k, i]_s - 1, \end{aligned}$$

so that, by (21) and (22),

$$(24) \quad \sum\nolimits_1 = \sum_{M|F}^{m>k} \lambda^{i+j}(M)|M|^{s+t+1}\{[m-k, i]_s[m-k, j]_t - p^{-n}[m-k, i]_s'[m-k, j]_t'\}.$$

The treatment of $\sum_2$ is much the same; we have

---

* For $s=i=0$, the symbol $[k, i]_s = k$.

$$\sum\nolimits_2 = |F|^{s+t} \sum_{\substack{(\alpha+\beta)F=\alpha AU+\beta BV \\ a<k,\, b=k}} \lambda^i(A)\lambda^j(B)\,|A|^{-s}\,|B|^{-t}$$

$$= |F|^{s+t} \sum_{M|F}^{m<k} \lambda^{i+j}(M)\,|M|^{-s-t} S_M,$$

where now

$$S_M = \sum_{\substack{(\alpha+\beta)FM^{-1}=\alpha AU+\beta BV \\ (A,B)=1;\, a<k-m=b}} \lambda^i(A)\lambda^j(B)\,|A|^{-s}\,|B|^{-t}$$

$$= |FM^{-1}| \sum_{\substack{(A,B)=1 \\ a<k-m=b}} \lambda^i(A)\lambda^j(B)\,|A|^{-s-1}\,|B|^{-t-1}$$

$$= |FM^{-1}| \sum_{a<k-m} (-1)^{ia+j(k-m)}\psi(a, k-m)p^{-n(s+1)a-n(t+1)(k-m)}$$

$$= |FM^{-1}|\,(-1)^{j(k-m)}p^{-nt(k-m)}\{[k-m, i]_s - p^{-n}[k-m, i]_s'\},$$

$[k, i]_s$ having the same meaning as in (23); therefore

$$(25)\qquad \begin{aligned} \sum\nolimits_2 = {} & \sum_{M|F}^{m>k} \lambda^{i+j}(M)\,|M|^{s+t+1}(-1)^{j(m-k)}p^{-nt(m-k)} \\ & \cdot\{[m-k, i]_s - p^{-n}[m-k, i]_s'\}. \end{aligned}$$

Similarly

$$(26)\qquad \begin{aligned} \sum\nolimits_3 = {} & \sum_{M|F}^{m>k} \lambda^{i+j}(M)\,|M|^{s+t+1}(-1)^{i(m-k)}p^{-ns(m-k)} \\ & \cdot\{[m-k, j]_t - p^{-n}[m-k, j]_t'\}. \end{aligned}$$

The sum $\sum_4$ is slightly different in that (A, B) may be of degree k; thus

$$\sum\nolimits_4 = |F|^{s+t} \sum_{M|F}^{m\leqq k} \lambda^{i+j}(M)\,|M|^{-s-t} S_M,$$

where

$$S_M = \sum_{\substack{(\alpha+\beta)FM^{-1}=\alpha AU+\beta BV \\ (A,B)=1;\, a=b=k-m}} \lambda^i(A)\lambda^j(B)\,|A|^{-s}\,|B|^{-t}$$

$$= |FM^{-1}| \sum_{\substack{(A,B)=1 \\ a=b=k-m}} \lambda^i(A)\lambda^j(B)\,|A|^{-s-1}\,|B|^{-t-1}$$

$$= |FM^{-1}|\,(-1)^{(i+j)(k-m)}\psi(k-m, k-m)p^{-n(s+t+2)(k-m)},$$

and therefore we have almost immediately

$$(27)\qquad \begin{aligned} \sum\nolimits_4 = {} & \sum_{M|F}^{m=k} \lambda^{i+j}(M)\,|M|^{s+t+1} \\ & + \sum_{M|F}^{m>k} \lambda^{i+j}(M)\,|M|^{s+t+1}(-1)^{(i+j)(m-k)}p^{-n(s+t)(m-k)}(1-p^{-n}). \end{aligned}$$

Substituting from (24), $\cdots$, (27) into (20), we find that the left member of (19) is

$$\sum_{M|F}^{m=k}\lambda^{i+j}(M)\,|\,M\,|^{s+t+1}+\sum_{M|F}^{m>k}\lambda^{i+j}(M)\,|\,M\,|^{s+t+1}\chi_m, \tag{28}$$

where

$$\begin{aligned}\chi_m = &\left\{1-(-1)^i p^{-ns}\right\}\left\{1-(-1)^j p^{-nt}\right\}\left\{[m-k,\,i]_s[m-k,\,j]_t\right.\\ &\left.-\,p^{-n}[m-k,\,i]_s'\,[m-k,\,j]_t'\right\}\\ &+\left\{1-(-1)^i p^{-ns}\right\}\left\{[m-k,i]_s-p^{-n}[m-k,i]_s'\right\}(-1)^{j(m-k)}p^{-nt(m-k)}\\ &+\left\{1-(-1)^j p^{-nt}\right\}\left\{[m-k,j]_t-p^{-n}[m-k,j]_t'\right\}(-1)^{i(m-k)}p^{-ns(m-k)}\\ &+(-1)^{i+j}p^{-n(s+t)(m-k)}(1-p^{-n})\\ =\ &1-(-1)^{i+j}p^{-n(s+t+1)},\end{aligned}$$

as may be verified without any calculation by applying (23) and then grouping the terms in an obvious way. This evidently completes the proof of (19) and therefore of Theorem 1.

We next prove a group of formulas that will be needed in §5 in deriving the expression sought for the number of solutions of (2).

THEOREM 2. *If F is of arbitrary degree, f; $2k$ is an even integer $>f$; α is any non-zero element of $GF(p^n)$; s, t two (real or complex) numbers; then*

$$\sum\rho_s(A)\rho_t(B) = p^{n(2k-f)(s+t+1)}\rho_{s+t+1}^k(F), \tag{29}$$

$$\sum\rho_s(A)\omega_t(B) = p^{n(2k-f)(s+t+1)}\omega_{s+t+1}^k(F), \tag{30}$$

$$\sum\omega_s(A)\omega_t(B) = p^{n(2k-f)(s+t+1)}\rho_{s+t+1}^k(F), \tag{31}$$

where, in each instance, the summation is extended over all polynomials A, B of degree $2k$, such that $\alpha F = A - B$; $\rho_t^k(F)$, $\omega_t(F)$ are defined by (6) and (7), respectively.

Exactly as in Theorem 1, the formulas (29), (30), (31) may be combined in a single relation involving the function $\Lambda_s^k(F, \lambda)$ defined by

$$\lambda^i(F)\Lambda_s^k(F,\lambda^i) = \left(1-\frac{(-1)^i}{p^{ns}}\right)\sum_{M|F}^{m>f-k}\lambda^i(M)\,|\,M\,|^s+\sum_{M|F}^{m=f-k}\lambda^i(M)\,|\,M\,|^s, \tag{32}$$

$\lambda(F)$ being defined by (16). The equations (18) may then be replaced by

$$\rho_s^k(F) = \Lambda_s^k(F, 1),\ \ \omega_s^k(F) = \Lambda_s^k(F, \lambda),$$

and the formulas (29), (30), (31) by

$$\sum_{\alpha F = A - B} \Lambda_s(A, \lambda^i)\Lambda_t(B, \lambda^j) = p^{n(2k-f)(s+t+1)}\Lambda^k_{s+t+1}(F, \lambda^{i+j}), \tag{33}$$

the summation being over all A, B of degree $2k$ for which $\alpha F = A - B$.

The proof of (33) is very similar to that of (19), except that wherever Lemma 2 is necessary, we now use case (II). It is scarcely necessary to give the proof in detail. We begin exactly as in (20), and we shall consider only the first sum, $\sum_1$; evidently

$$\begin{aligned}
\sum\nolimits_1 &= \sum_{\substack{\alpha F = AU - BV \\ a, b < k}} \lambda^i(U)\lambda^j(V) \mid U \mid^s \mid V \mid^t \\
&= p^{2nk(s+t)} \sum_{\substack{\alpha F = AU - BV \\ a, b < k}} \lambda^i(A)\lambda^j(B) \mid A \mid^{-s} \mid B \mid^{-t} \\
&= p^{2nk(s+t)} \sum_{M \mid F}^{m<k} \lambda^{i+j}(M) \mid M \mid^{-s-t} S_M,
\end{aligned}$$

where, exactly as in (21),

$$\begin{aligned}
S_M &= \sum_{\substack{(A, B) = 1 \\ a, b < k-m}} \lambda^i(A)\lambda^j(B) \mid A \mid^{-s} \mid B \mid^{-t} \sum_{\substack{\alpha F M^{-1} = AU - BV \\ a+u = b+v = 2k-m}} 1 \\
&= p^{2nk} \mid M \mid^{-1} \sum_{\substack{(A, B) = 1 \\ a, b < k-m}} \lambda^i(A)\lambda^j(B) \mid A \mid^{-s-1} \mid B \mid^{-t-1}
\end{aligned}$$

by case (II) of Lemma 2. Therefore

$$S_M = p^{2nk} \mid M \mid^{-1} \sum_{a, b < k-m} (-1)^{ia+jb} \psi(a, b) p^{-n(s+1)a - n(t+1)b},$$

which may be evaluated by following the method applied to (22). Thus we find that

$$\begin{aligned}
\sum\nolimits_1 &= p^{2nk(s+t+1)} \sum_{M \mid F}^{m<k} \lambda^{i+j}(M) \mid M \mid^{-s-t-1} \\
&\quad \cdot \{ [k - m, i]_s [k - m, j]_t - p^{-n} [k - m, i]_s' [k - m, j]_t' \} \\
&= \lambda^{i+j}(F) p^{n(2k-f)(s+t+1)} \sum_{M \mid F}^{m > f-k} \lambda^{i+j}(M) \mid M \mid^{s+t+1} \\
&\quad \cdot \{ [m + k - f, i]_s [m + k - f, j]_t - p^{-n} [m + k - f, i]_s' [m + k - f, j]_t' \}.
\end{aligned} \tag{34}$$

Similarly, we find that

$$\begin{aligned}
\sum\nolimits_2 &= \lambda^{i+j}(F) p^{n(2k-f)(s+t+1)} \sum_{M \mid F}^{m > f-k} \lambda^{i+j}(M) \mid M \mid^{s+t+1} \\
&\quad \cdot (-1)^{j(m+k-f)} p^{-nt(m+k-f)} \{ [m + k - f, i]_s - p^{-n} [m + k - f, j]_s' \};
\end{aligned} \tag{35}$$

$$\sum\nolimits_3 = \lambda^{i+j}(F)p^{n(2k-f)(s+t+1)} \sum_{M|F}^{m>f-k} \lambda^{i+j}(M) \mid M \mid^{s+t+1} \cdot (-1)^{i(m+k-f)} p^{-ns(m+k-f)} \{ [m+k-f,j]_t - p^{-n}[m+k-f,j]_t' \}; \tag{36}$$

$$\begin{aligned} \sum\nolimits_4 = \lambda^{i+j}(F)p^{n(2k-f)(s+t+1)} \Bigg\{ & \sum_{M|F}^{m=f-k} \lambda^{i+j}(M) \mid M \mid^{s+t+1} \\ & + (1-p^{-n}) \sum_{M|F}^{m>f-k} \lambda^{i+j}(M) \mid M \mid^{s+t+1} \\ & \cdot (-1)^{(i+j)(m+k-f)} p^{-n(s+t)(m+k-f)} \Bigg\}. \end{aligned} \tag{37}$$

Combining (34), $\cdots$, (37) exactly as in (28) (the corresponding point in the proof of Theorem 1) we complete the proof of (33) and therefore of Theorem 2.

4. **Number of solutions of** (1). We begin with the case $s=1$ and then proceed by induction to the formulas (3) and (4) for general s.

THEOREM 3. *If α is an element of $GF(p^n), \neq 0$ or 1; F is of even degree, $2k$; then the number of solutions of*

$$(1-\alpha)F = X^2 - \alpha Y^2 \tag{38}$$

in (primary) X, Y of degree k, is

(I) $\displaystyle\sum_{M|F}^{m=k} 1$ *for α a square in $GF(p^n)$,*

(II) $\displaystyle\sum_{M|F} (-1)^m$ *for α a non-square in $GF(p^n)$.*

The case (I) is almost trivial. Let $\alpha = \beta^2$, β in $GF(p^n)$ and $\neq \pm 1$, so that (38) becomes

$$F = \frac{X+\beta Y}{1+\beta} \frac{X-\beta Y}{1-\beta} = UV,$$

say. Evidently U and V are primary of degree k. But the number of solutions of $F=UV$, U and V of equal degree, is of course the number of divisors of F that are of degree k. Since U, V uniquely determines X, Y, this establishes the formula (I).

(II)* α is now not the square of any element of $GF(p^n)$; however it is a square in the Galois field of order p^{2n}, $GF(p^{2n})$, which contains the original

* This case can be deduced from the general theory of quadratic fields over $\mathfrak{D}$, worked out in detail by Artin, Mathematische Zeitschrift, vol. 19 (1924), pp. 153–246. However we shall make no use of this theory here.

$GF(p^n)$. Put $\alpha=\theta^2$, so that θ is in $GF(p^{2n})$ but not in $GF(p^n)$; in particular $\theta\neq\pm 1$. Then as above

$$F = \frac{X+\theta Y}{1+\theta}\,\frac{X-\theta Y}{1-\theta} = UV,$$

U and V now being over $GF(p^{2n})$ and of equal degree. Put

$$U = A + \theta B,\ V = A' + \theta B',$$

where A, B, A', B' are all over $GF(p^n)$; A and A' are primary and of degree k; B and B' are of degree less than k and not necessarily primary. Then

$$X + \theta Y = (1+\theta)(A+\theta B),$$
$$X - \theta Y = (1-\theta)(A'+\theta B'),$$

whence $A=A'$, $B=-B'$. Therefore we seek the number of solutions of

$$F = (A+\theta B)(A-\theta B), \tag{39}$$

where A is primary of degree k, and B is of lesser degree and need not be primary. This can be determined readily if we make use of two well known properties of polynomials over a Galois field: first, *an irreducible polynomial over $GF(p^n)$ factors in $GF(p^{2n})$ if and only if its degree is even*; second, *a polynomial over $GF(p^n)$ can be expressed as a product of irreducible polynomials over $GF(p^n)$ in essentially one way.*

Suppose now $F=Q^l$, Q irreducible of degree q. Clearly if l and q are both odd, there are no factorizations (39); if q is odd but l is even, there is one such factorization. However if q is even, there are $l+1$ factorizations. In other words the number of solutions of (39) in this case is

$$1 + (-1)^q + \cdots + (-1)^{lq}.$$

Similarly if $F=\Pi Q^l$, Q irreducible, the number of solutions of (39) is

$$\prod_{Q|F}\{1+(-1)^q+\cdots+(-1)^{ql}\} = \sum_{M|F}(-1)^m.$$

This completes the proof of formula (II).

We are now able to prove our first principal result.

THEOREM 4. *If $\alpha_1,\cdots,\alpha_s,\beta_1,\cdots,\beta_s$ are non-zero elements of $GF(p^n)$, such that*

$$\gamma_i = \alpha_i+\beta_i \neq 0,$$
$$\gamma = \gamma_1+\cdots+\gamma_s \neq 0;$$

F is of even degree, $2k$; then the number of solutions of (1) *is $\rho_{s-1}(F)$ if*

$$(-1)^s\alpha_1\cdots\alpha_s\beta_1\cdots\beta_s \tag{40}$$

is a square in $GF(p^n)$; and is $\omega_{s-1}(F)$ if (40) *is a non-square in $GF(p^n)$.*

The case $s=1$ of this theorem is clearly true by virtue of Theorem 3 and the definition of $\rho_0(F)$ and $\omega_0(F)$. Assume the theorem true for all values up to and including s. In order to effect the induction it is necessary to consider two cases: (I) for some $j \leqq s+1$,

$$\gamma^{(s+1)} = \sum_{i=1}^{s+1} \gamma_i \neq \gamma_j; \tag{41}$$

(II) for no j is (41) satisfied.

(I) Assume the notation is such that $\gamma^{(s)} = \gamma_1 + \cdots + \gamma_s \neq 0$. By hypothesis $\gamma_{s+1} \neq 0$, $\gamma^{(s+1)} \neq 0$. If we put

$$\begin{gathered} \gamma^{(s+1)}(F) = \gamma^{(s)}A + \gamma_{s+1}B, \\ |A| = |B| = |F|, \end{gathered} \tag{42}$$

then, since our theorem is assumed true for s, it is obvious that the number of solutions in question for $s+1$ is

$$\begin{array}{llll} \text{(i)} & \sum \rho_{s-1}(A)\rho_0(B), & \text{(ii)} & \sum \rho_{s-1}(A)\omega_0(B), \\ \text{(iii)} & \sum \omega_{s-1}(A)\rho_0(B), & \text{(iv)} & \sum \omega_{s-1}(A)\omega_0(B), \end{array} \tag{43}$$

according as

(i) (5) and $-\alpha_{s+1}\beta_{s+1}$ are both squares,
(ii) (5) is a square, $-\alpha_{s+1}\beta_{s+1}$ a non-square,
(iii) (5) a non-square, $-\alpha_{s+1}\beta_{s+1}$ a square,
(iv) (5) and $-\alpha_{s+1}\beta_{s+1}$ both non-squares;

the sums (43) being taken over all A, B satisfying (42). If now we apply Theorem 1, it is clear that the induction is complete for case (I).

(II) Since (41) is satisfied for no j, it is clear that $\gamma_1 = \gamma_2 = \cdots = \gamma_{s+1}$, and therefore s is a multiple of p. As a consequence of this,

$$\gamma^{(s-1)} = \gamma_1 + \cdots + \gamma_{s-1} \neq 0, \ \gamma_s + \gamma_{s+1} \neq 0.$$

Let us now put, in place of (42),

$$\gamma^{(s+1)}F = \gamma^{(s-1)}A + (\gamma_s + \gamma_{s+1})B^*, \ |A| = |B| = |F|; \tag{44}$$

in place of (43) we now have

$$\sum \rho_{s-2}(A)\rho_1(B), \ \sum \rho_{s-2}(A)\omega_1(B), \text{ etc.},$$

summed over all A, B satisfying (42). The induction is completed as in case (I).

---

* Or $F = -A + 2B$.

COROLLARY. *Let F be of degree $2k$ over $GF(p^n)$, s not a multiple of p. Then the number of solutions of*

$$2sF = X_1^2 + X_2^2 + \cdots + X_{2s}^2$$

in (primary) X_i of degree k is $\rho_{s-1}(F)$ if

(i) *s is even,*

(ii) *s is odd, n is even,*

(iii) *s and n are odd,* $p\equiv 1 \pmod 4$;

the number of solutions is $\omega_{s-1}(F)$ otherwise, that is, if

(iv) *s and n are odd,* $p\equiv 3 \pmod 4$.

5. **Number of solutions of (2).** Our second principal result is contained in the following theorem.

THEOREM 5. *If $\alpha, \alpha_1, \cdots, \alpha_s, \beta_1, \cdots, \beta_s$ are non-zero elements of $GF(p^n)$, such that*

$$\gamma_i = \alpha_i + \beta_i \neq 0,\ \gamma = \gamma_1 + \cdots + \gamma_s = 0;$$

F is of arbitrary degree, f; $2k$ is an even integer $>f$; then the number of solutions of (2) is

$$p^{n(s-1)(2k-f)}\overset{k}{\rho}_{s-1}(F) \quad or \quad p^{n(s-1)(2k-f)}\overset{k}{\omega}_{s-1}(F)$$

according as (40) is or is not a square in $GF(p^n)$.

Take first $s=1$; we may write (2) in the form

$$\alpha F = X^2 - Y^2,\ \deg X = \deg Y = k \tag{45}$$

(so that (40) is necessarily a square). But (45) is equivalent to

$$F = UV,\ \deg U = k,\ \deg V = f - k;^*$$

therefore the number of solutions of (45) is the number of divisors of F of degree $f-k$, i.e.

$$\sum_{M|F}^{m=f-k} 1 = \rho_0^k(F).$$

Since (40) is necessarily a square, our theorem holds for $s=1$.

For $s>1$, we make use of Theorem 4. Since $\gamma=0, \gamma_1\neq 0$, plainly $\gamma-\gamma_1\neq 0$. Let us put

$$\alpha F = \gamma_1 A + (\gamma - \gamma_1)B,\ \deg A = \deg B = 2k. \tag{46}$$

---

* U and V are of course primary.

By Theorem 4 we may express the number of solutions of

$$\gamma_1 A = \alpha_1 X_1^2 + \beta_1 Y_1^2, \quad (\gamma - \gamma_1)B = \alpha_2 X_2^2 + \cdots + \beta_s Y_s^2,$$

in terms of $\rho_0(A)$, $\omega_0(A)$; $\rho_{s-2}(A)$, $\omega_{s-2}(A)$, respectively. Thus, if $-\alpha_1\beta_1$ and $(-1)^{s-1}\alpha_2 \cdots \beta_s$ are both squares, the number of solutions of (2) is

$$\sum \rho_0(A)\rho_{s-2}(B)$$

summed over all A, B satisfying (46). Applying Theorem 2, this sum is

$$p^{n(s-1)(2k-f)}\overset{k}{\rho}_{s-1}(F),$$

which proves the theorem in this case. The proof is exactly the same in each of the remaining three cases and need not be repeated. This completes the proof of Theorem 5.

COROLLARY 1. *If all the hypotheses of Theorem 5 are true, and in addition $k>f$, then the number of solutions of (2) is*

$$\left(1 - \frac{1}{p^{n(s-1)}}\right)p^{n(s-1)(2k-f)} \sum_{M|F} |M|^{s-1}$$

or

$$\left(1 - \frac{1}{p^{n(s-1)}}\right)p^{n(s-1)(2k-f)} \sum_{M|F} (-1)^{f-m} |M|^{s-1}$$

according as (40) is or is not a square in $GF(p^n)$, the summations now being taken over all M dividing F.

COROLLARY 2.* *Let F be of degree f over $GF(p^n)$, $2k$ an even integer $>f$, s a multiple of p, $\alpha \neq 0$. Then the number of solutions of*

$$\alpha F = X_1^2 + \cdots + X_{2s}^2$$

in (primary) X_i of degree k is

$$p^{n(s-1)(2k-f)}\overset{k}{\rho}_{s-1}(F)$$

if $ns(p-1)/2$ is even; the number of solutions is

$$p^{n(s-1)(2k-f)}\overset{k}{\omega}_{s-1}(F)$$

if $ns(p-1)/2$ is odd.

---

* Cf. the corollary to Theorem 4.

DUKE UNIVERSITY,
DURHAM, N.C.

ON THE CLASS NUMBER OF A CYCLIC FIELD*

BY

CLAIBORNE G. LATIMER

1. Introduction. Let Ω be the field defined by a primitive mth root of unity, m an integer >2, and let F be a subfield of Ω. In a recent article,† Gut showed that if F is real, the class number may be written $h=\delta/R$, where R is the regulator of F and δ is a product involving certain group characters. If F is imaginary, he showed that $h=h_1\cdot h_2$, where h_1 is a closed expression and $h_2=\delta/R$, δ and R being as before. If $F=\Omega$ and m is an odd prime, Gut's h_1 and h_2 are the same, except perhaps for sign, as Kummer's well known first and second factors of the class number.

We shall assume hereafter that the Galois group $\mathfrak{A}$ of F is cyclic. In this case, as noted by Gut, the δ in his expression for h, or h_2, may be written as a determinant. Employing this determinantal form, we shall show that δ/R, and hence h or h_2, is equal to $N(\tau)/N(\mathfrak{R})$, where $N(\mathfrak{R})$ is the norm of a non-singular ideal $\mathfrak{R}$, in a set $\mathfrak{G}$ of elements in a certain commutative algebra, and $N(\tau)$ is the norm of a principal ideal $\{\tau\}$ in $\mathfrak{G}$, τ being an element in $\mathfrak{R}$.‡

In certain cases our results may be expressed in terms of an ideal in a cyclotomic field. (See Theorem 2.) For the case where F is a cubic field, the discriminant of which is the square of a prime, Theorem 2 is equivalent to Eisenstein's result that the number of classes of certain "associated (cubic) forms" is $h=\mu^2-\mu\nu+\nu^2$, where μ, ν are rational integers.§

2. The ratio of two determinants. Let F be of degree E and let s be a generating substitution of $\mathfrak{A}$. If θ is a number of F, not rational, it will be understood that $\theta^{(i)}\equiv s^i(\theta)$ $(i=1, 2, \cdots, E)$, $\theta^{(E)}=\theta^{(0)}=\theta$. Let $e\equiv E$ or $e\equiv E/2$ according as F is real or imaginary. Then $\theta^{(i+e)}$ is the conjugate imaginary of $\theta^{(i)}$ $(i=0, 1, 2, \cdots, e-1)$.

Let $\eta_1, \eta_2, \cdots, \eta_n$ be a fundamental set of units of F. By Dirichlet's well known theorem, $n=e-1$. Since every η_i' belongs to F,

---

* Presented to the Society, December 28, 1931; received by the editors August 27, 1932.

† *Die Zetafunktion, die Klassenzahl und die Kronecker'sche Grenzformel eines beliebigen Kreiskörpers*, Commentarii Mathematici Helvetici, vol. 1 (1929), p. 160.

‡ It will be understood that we use the same definitions of terms referring to ideals in $\mathfrak{G}$ as are given by MacDuffee in his article *An introduction to the theory of ideals*, etc., these Transactions, vol. 31 (1929), p. 71. In case $\mathfrak{G}$ is a set of integral algebraic numbers, these definitions are equivalent to the usual definitions.

§ Journal für Mathematik, vol. 29 (1845), p. 49.

(1) $$\eta_i' = u_i\eta_1^{\alpha_{i1}}\eta_2^{\alpha_{i2}}\cdots\eta_n^{\alpha_{in}} \qquad (i = 1, 2, \cdots, n),$$

where u_i is a root of unity and the α's are rational integers. Let the nth order matrix $A \equiv (\alpha_{ij})$ and let I be the identity matrix.

LEMMA 1. *A is a root of*

(2) $$f(x) \equiv x^n + x^{n-1} + \cdots + x + I = 0,$$

and it is not a root of an equation of lower degree with rational coefficients.

By (1), if $0 \leqq k < E$,

$$\eta_i^{(k)} = u_i^{(k)}\eta_1^{\alpha_{i1}^{(k)}}\eta_2^{\alpha_{i2}^{(k)}}\cdots\eta_n^{\alpha_{in}^{(k)}} \qquad (i = 1, 2, \cdots, n),$$

where $u_i^{(k)}$ is a root of unity and the matrix $(\alpha_{ij}^{(k)}) = A^k$. Since $\eta_i\eta_i' \cdots \eta_i^{(E-1)} = \pm 1$, it follows that A is a root of

$$f_1(x) \equiv x^{E-1} + x^{E-2} + \cdots + x + I = 0.$$

If F is real, it follows that A is a root of (2). Suppose F is imaginary. Then

$$f_1(A) = f(A)(A^e + I) = 0.$$

To prove that A is a root of (2), it suffices to show that $A^e + I$ is non-singular.

Let $A^e + I \equiv (\beta_{ij})$. We have

$$\eta_i\eta_i^{(e)} = v_i\eta_1^{\beta_{i1}}\eta_2^{\beta_{in}}\cdots\eta_n^{\beta_{i2}} \qquad (i = 1, 2, \cdots, n),$$

where v_i is a root of unity. Suppose (β_{ij}) is singular. Then the system of equations

$$\sum_{j=1}^{n}\beta_{ji}x_j = 0 \qquad (i = 1, 2, \cdots, n)$$

has a solution in rational integers, not all zero, and

$$\phi \equiv \prod_{i=1}^{n}(\eta_i\eta_i^{(e)})^{x_i}$$

and every $\phi^{(i)}$ is a root of unity. Let $\lg\theta$ be the real logarithm of $|\theta|$. Then $\lg\theta^{(e)} = \lg\theta$ and, since $|\phi^{(i)}| = 1$,

$$\sum_{j=1}^{n}x_j \lg\eta_j^{(i)} = 0 \qquad (i = 0, 1, 2, \cdots, n-1).$$

From this it follows that the regulator, $R = \pm|\lg\eta_i \;\lg\eta_i' \cdots \lg\eta_i^{(n-1)}|$ $(i = 1, 2, \cdots, n)$, of F is zero.* But this is known to be false. Hence $A^e + I$ is non-singular and A is a root of (2).

* We take the same definition of R as that used by Gut, loc. cit., p. 200.

It may be shown by the same method employed by Pollaczek on a similar problem*, that A is not a root of an equation of degree $<n$ with rational coefficients. The lemma follows.

Let $x_1, x_2, \cdots, x_n$ be independent variables and let

$$x_i^{(k)} \equiv \alpha_{1i}^{(k)} x_1 + \alpha_{2i}^{(k)} x_2 + \cdots + \alpha_{ni}^{(k)} x_n \quad (i = 1, 2, \cdots, n). \tag{3}$$

For a fixed k, the matrix of the forms $x_i^{(k)}$ is the transpose of A^k. By Lemma 1, $A^e = I$. Thus we have a cyclic group of linear homogeneous substitutions $S, S^2, \cdots, S^e = 1$, on the x's. For every pair of integers i, k,

$$S^k(x_1^{(i)}, x_2^{(i)}, \cdots, x_n^{(i)}) = (x_1^{(i+k)}, x_2^{(i+k)}, \cdots, x_n^{(i+k)}), \tag{4}$$

it being understood that if $j \equiv j_1 \pmod{e}$, $0 \leqq j_1 < e$, then $x_i^{(j)} = x_i^{(j_1)}$, $x_i^{(0)} = x_i$ $(i = 1, 2, \cdots, n)$.

If θ is a unit of F, by (1) and (3)

$$\begin{aligned} \theta &= u_1 \eta_1^{x_1} \eta_2^{x_2} \cdots \eta_n^{x_n}, \\ \theta' &= u_2 \eta_1^{x_1'} \eta_2^{x_2'} \cdots \eta_n^{x_n'}, \\ &\vdots \\ \theta^{(n)} &= u_n \eta_1^{x_1^{(n)}} \eta_2^{x_2^{(n)}} \cdots \eta_n^{x_n^{(n)}}, \end{aligned} \tag{5}$$

where the u's are roots of unity and the x's are rational integers. It will be observed that if we apply a substitution s^i to θ, the resulting unit is the same, except perhaps for a factor which is a root of unity, as that obtained by applying the substitution S^i to the x's when θ is written as in the first equation above.

If $0 \leqq t < e$ and if $i + k \equiv t \pmod{e}$, by (4) and (5),

$$\theta^{(t)} = u \eta_1^{(i)\, z_1} \eta_2^{(i)\, z_2} \cdots \eta_n^{(i)\, z_n},$$

where u is a root of unity and $z_j = x_j^{(k)}$ $(j = 1, 2, \cdots, n)$. Let the determinant of the x's in the first n equations of (5) be $\Psi(x_1, x_2, \cdots, x_n)$ and let

$$\delta(\theta) \equiv \begin{vmatrix} \lg \theta & \lg \theta' & \cdots \lg \theta^{(n-1)} \\ \lg \theta' & \lg \theta'' & \cdots \lg \theta^{(n)} \\ \cdot & \cdot & \cdots \quad \cdot \\ \lg \theta^{(n-1)} & \lg \theta^{(n)} & \cdots \lg \theta^{(n-3)} \end{vmatrix}. \tag{6}$$

* Mathematische Zeitschrift, vol. 21 (1924), pp. 8, 9; Bulletin of the National Research Council, No. 62, *Algebraic Numbers*, II, pp. 94–96.

Employing (6) and the same rule for the multiplication of determinants as for matrices, we find $\Psi(x_1, x_2, \cdots, x_n)\cdot R = \pm\delta(\theta)$. Hence*

$$\Psi(x_1, x_2, \cdots, x_n) = \pm \frac{\delta(\theta)}{R}\cdot \tag{7}$$

3. **The set $\mathfrak{G}$.** The algebraic roots of (2) are distinct and hence the same is true of the factors of $f(x)$ which are irreducible in the rational field.

Let C be any matrix such that (2) is the equation of minimum degree with rational coefficients, the leading coefficient being unity, which has C as a root. Let $\mathfrak{G}$ be the set of all polynomials in C with rational integral coefficients. It has been shown that there is a one-to-one correspondence between the classes of ideals in $\mathfrak{G}$ and certain classes of matrices.† Since (2) is the minimum equation of A, by the proof of this result, there is a non-singular ideal $\mathfrak{K}$, in $\mathfrak{G}$, with a basis $\omega_1, \omega_2, \cdots, \omega_n$ such that

$$C\omega_i = \alpha_{i1}\omega_1 + \alpha_{i2}\omega_2 + \cdots + \alpha_{in}\omega_n \qquad (i = 1, 2, \cdots, n).$$

Let τ be an element in $\mathfrak{K}$. By the last equations and (3),

$$\begin{aligned}
\tau &= x_1\omega_1 + x_2\omega_2 + \cdots + x_n\omega_n,\\
C\tau &= x_1'\omega_1 + x_2'\omega_2 + \cdots + x_n'\omega_n,\\
&\cdots\cdots\cdots\cdots\cdots\cdots\\
C^{n-1}\tau &= x_1^{(n-1)}\omega_1 + x_2^{(n-1)}\omega_2 + \cdots + x_n^{(n-1)}\omega_n,
\end{aligned}$$

where the x's are rational integers. The determinant of the coefficients of the ω's is $\Psi(x_1, x_2, \cdots, x_n)$.

$I, C, C^2, \cdots, C^{n-1}$ form a basis of $\mathfrak{G}$. Hence

$$\omega_i = g_{i1}I + g_{i2}C + \cdots + g_{in}C^{n-1} \qquad (i = 1, 2, \cdots, n),$$

where the g's are rational integers such that the absolute value of the determinant $|g_{ij}|$ is the norm of $\mathfrak{K}$. If we employ the last equations to eliminate the ω's in the above expressions for $C^{i-1}\tau (i=1, 2, \cdots, n)$, in the resulting equations the determinant of the coefficients of the powers of C is $\Psi(x_1, x_2, \cdots, x_n)\cdot N(\mathfrak{K})$. But the $C^{i-1}\tau$ form a basis of the principal ideal $\{\tau\}$. Hence $\Psi(x_1, x_2, \cdots, x_n)\cdot N(\mathfrak{K}) = \pm N(\tau)$. Since $\mathfrak{K}$ is non-singular, $N(\mathfrak{K})\neq 0$.

Therefore by (7) we have

---

* By employing Lemma 1, it may be shown that Ψ is an invariant of the above-mentioned substitution group. See Fricke, *Lehrbuch der Algebra*, vol. 2, p. 14.

† Latimer and MacDuffee, *A correspondence between classes of ideals and classes of matrices*, Annals of Mathematics, vol. 34 (1933).

LEMMA 2. *If $\theta = u\eta_1^{x_1}\eta_2^{x_2}\cdots\eta_n^{x_n}$ is a unit of F, where u is a root of unity, then*

$$\pm\frac{\delta(\theta)}{R} = \pm\Psi(x_1, x_2, \cdots, x_n) = \frac{N(\tau)}{N(\mathfrak{R})},$$

where $\mathfrak{R}$ is a non-singular ideal in $\mathfrak{G}$, with a basis $\omega_1, \omega_2, \cdots, \omega_n$ such that

$$C\omega_i = \alpha_{i1}\omega_1 + \alpha_{i2}\omega_2 + \cdots + \alpha_{in}\omega_n \qquad (i = 1, 2, \cdots, n)$$

and $N(\tau)$ is the norm of the principal ideal $\{\tau\}$, $\tau = x_1\omega_1 + x_2\omega_2 + \cdots + x_n\omega_n$.

4. **Proof of principal theorem.** Let $\mathfrak{N}$ be the group which has as its elements the $\phi(m)$ integers in a reduced set of residues, modulo m. The numbers of F are those numbers of Ω which are unaltered under every substitution (ρ, ρ^a), where ρ is a primitive mth root of unity and a is an integer in a subgroup $\mathfrak{U}$ of $\mathfrak{N}$. Let the co-sets (Nebengruppen) of $\mathfrak{N}$ with respect to $\mathfrak{U}$ be $\mathfrak{U}_0 = \mathfrak{U}, \mathfrak{U}_1, \mathfrak{U}_2, \cdots, \mathfrak{U}_{E-1}$. Then $\mathfrak{U}_i = \gamma_i\mathfrak{U}$ where the γ_i are properly chosen integers. The factor group $\mathfrak{N}/\mathfrak{U}$ is simply isomorphic with $\mathfrak{A}$,* which by hypothesis is cyclic. Hence we may assume that $s = (\rho, \rho^\gamma)$, where γ is an integer such that $\gamma^E \equiv a \pmod{m}$, a an element in $\mathfrak{U}$. If m is odd, we may assume that γ is odd, while if m is even, γ is necessarily odd since the same is true of a.

If F is real, by Gut's results, $h = \Delta/R$ where†

$$\Delta = \prod_{\chi}\sum_{k=1}^{m/2} - \chi(k)\log\sin\frac{\pi k}{m}\cdot \tag{8}$$

In the product, χ ranges over all the elements, except the identity element, of a group of characters which is simply isomorphic with $\mathfrak{A}$. Since $\mathfrak{A}$ is cyclic, we have

$$\Delta = \prod_{t=1}^{n}\sum_{k=1}^{m/2} - \chi^t(k)\log\sin\frac{\pi k}{m}, \tag{9}$$

where χ is a fixed character. It may be shown that if a and b are prime to m, $\chi(a) = \chi(b)$ if and only if a and b are congruent, modulo m, to elements in the same co-set $\mathfrak{U}_i$. After proper choice of notation, we may assume that if a belongs to $\mathfrak{U}_i$, $\chi(a) = \zeta^i$, where ζ is a primitive eth root of unity. Employing $\chi(m-k) = \chi(k)$, $\chi(k) = 0$ if $(m, k) > 1$, and $\sum_{k=1}^{m-1}\chi^t(k) = 0$ $(0 < t < e)$, it may be shown that

$$2\sum_{k=1}^{m/2}\chi^t(k)\log\sin\frac{\pi k}{m} = \sum_{k=1}^{m-1}\chi^t(k)\lg(1-\rho^k) = \sum_{i=0}^{n}\zeta^{ti}\lg\lambda_i, \tag{10}$$

* Weber, *Lehrbuch der Algebra*, 2d edition, vol. 2, p. 75.

† Gut, loc. cit., pp. 200, 223.

where $\lambda_0 = \Pi(1-\rho^a)$, a ranging over all the elements of $\mathfrak{U}$, and $\lambda_i = \lambda_0^{(i)}$ $(i=1, 2, \cdots, n)$. Employing a well known property of cyclic determinants, it may be shown from (9) and (10) that

$$\Delta = 2^{-n} \prod_{t=1}^{n} \sum_{i=0}^{n} \zeta^{ti} \lg \lambda_i = \pm \delta(\theta),$$

where $\theta = (\lambda_1/\lambda_0)^{1/2}$.* Hence $h = \pm\delta(\theta)/R$. We shall show that θ is a unit of F. Since F is real, $\mathfrak{U}$ contains -1. Therefore θ is a product of units in the form

$$\left[\frac{(1-\rho^{\gamma a})(1-\rho^{-\gamma a})}{(1-\rho^{a})(1-\rho^{-a})}\right]^{1/2} = \pm \rho^{(1-\gamma)a/2}\left(\frac{1-\rho^{\gamma a}}{1-\rho^{a}}\right).$$

Since γ is odd, the unit on the left belongs to Ω, and hence the same is true of θ. Since θ is unaltered under every substitution (ρ, ρ^a), a in $\mathfrak{U}$, it belongs to F.

If F is imaginary, Gut's expression for h may be written $h = h_1 \cdot h_2$, where h_1 is a closed expression and $h_2 = \Delta/R$, where Δ is exactly the same as the right side of (8), except that in this case χ ranges over those characters, except the principal character, such that $\chi(-1) = 1$.† The whole group of characters is simply isomorphic with $\mathfrak{A}$ and hence every character is a power of one of them. For a generating character χ, we have $\chi(-1) = -1$. Hence $h_2 = \Delta/R$, where

$$\Delta = \prod_{t=1}^{n} \sum_{k=1}^{m/2} - \chi^{2t}(k) \log \sin \frac{\pi k}{m} \cdot$$

Since $s^e(\theta)$ is the conjugate imaginary of θ, the co-set $\mathfrak{U}_e$ contains -1 and we may take as the elements of $\mathfrak{U}_{i+e}$ the negatives of the elements in the corresponding $\mathfrak{U}_i$. If a and b are prime to m and a is in $\mathfrak{U}_i$, then $\chi^2(a) = \chi^2(b)$ if and only if b is congruent to an element in $\mathfrak{U}_i$ or in $\mathfrak{U}_{i+e}$. The notation for the co-sets may be so chosen that if a belongs to $\mathfrak{U}_i$ then $\chi^2(a) = \zeta^i$, where ζ is a primitive eth root of unity. If we define the λ_i as before, let $\theta \equiv (\lambda_1 \cdot \lambda_{e+1} / \lambda_0 \cdot \lambda_e)$ and employ the fact that λ_{i+e} is the conjugate imaginary of λ_i, we find as before that $\Delta = \pm\delta(\theta)$, $h_2 = \pm\delta(\theta)/R$ and θ is a real unit of F. By Lemma 2, we have then the following, except the last sentence.

THEOREM 1. *Let F be a field, of degree E, which is cyclic with respect to the rational field. Let $e = E$ or $e = E/2$ according as F is real or imaginary, and let*

---

* For a special case of this, see Fueter, *Die Klassenzahl zyklischer Körper*, etc., Journal für Mathematik, vol. 147 (1917), p. 183.

† Gut, loc. cit., pp. 201, 223.

$n=e-1$. *Let* $\mathfrak{G}$ *be the set of all polynomials with rational integral coefficients in the nth order matrix* $A=(\alpha_{ij})$, *where the* α*'s are given in* (1). *If* F *is real let* H *be the class number of* F, *and if* F *is imaginary let* H *be the absolute value of Gut's second factor of the class number. Then*

$$H = N(\tau)/N(\mathfrak{K}),$$

where $N(\mathfrak{K})$ *is the norm of a non-singular ideal* $\mathfrak{K}$ *in* $\mathfrak{G}$ *and* $N(\tau)$ *is the norm of a principal ideal* $\{\tau\}$ *in* $\mathfrak{G}$, τ *being an element in* $\mathfrak{K}$. *If* F *is the field defined by a primitive mth root of unity, m an odd prime,* $\pm H$ *is Kummer's second factor of the class number.*

To prove the last sentence of the theorem, it suffices to note that our θ, $\delta(\theta)$, R, when properly specialized, are identical, except perhaps for sign, with Kummer's $e(\alpha)$, D, Δ respectively.*

It will be observed that by the proof of the above theorem, $\pm H$ is represented by the form $\Psi(x_1, x_2, \cdots, x_n)$, which, as previously noted, is an invariant of the cyclic substitution group defined by the transpose of A.

5. **A special case of Theorem 1.** Suppose e of Theorem 1 is an odd prime, F real or imaginary. Let ζ be a primitive eth root of unity. ζ is a root of (2), and $1, \zeta, \zeta^2, \cdots, \zeta^{n-1}$ form a basis of the integral numbers in the field K defined by ζ. Hence by Lemma 1, $\mathfrak{G}$ is equivalent to the set of all integral algebraic numbers in K. Then, by well known theorems in algebraic numbers, there is an ideal $\mathfrak{L}$ such that $\{\tau\}=\mathfrak{K}\mathfrak{L}$ and $N(\tau)=N(\mathfrak{K})\cdot N(\mathfrak{L})$. We have then

THEOREM 2. *If e in Theorem* 1 *is an odd prime,*

$$H = N(\mathfrak{L}),$$

where $\mathfrak{L}$ *is an ideal in the field defined by a primitive eth root of unity.*

If F is the field defined by a primitive mth root of unity, m an odd prime, and if $e=(m-1)/2$ is also an odd prime, it may be shown that Kummer's first factor of the class number is the norm of a principal ideal in K, K as above. Hence the class number of F is the norm of an ideal in K.

* Journal für Mathematik, vol. 40 (1850), pp. 110, 99; Bulletin of the National Research Council, loc. cit., p. 34.

UNIVERSITY OF KENTUCKY,
LEXINGTON, KY.

THE BOUNDARY VALUES OF ANALYTIC FUNCTIONS. II*

BY

JOSEPH L. DOOB

Let

$$f_1(z), f_2(z), \cdots \tag{0.1}$$

be a uniformly bounded sequence of functions analytic for $|z|<1$. By a theorem of Fatou,† $\lim_{r\to 1} f_n(re^{it})$ exists almost everywhere on the interval $0 \leq t < 2\pi$, defining a boundary function $F_n(e^{it}) = \lim_{r\to 1} f_n(re^{it})$ almost everywhere on $|z|=1$, $z=e^{it}$. A new sequence

$$F_1(z), F_2(z), \cdots \tag{0.2}$$

is thus determined. What are the relations between these two sequences? More generally, let the sequence (0.1) consist of functions meromorphic for $|z|<1$. In §2 below, a boundary function $\mathcal{J}_n(z)$ will be defined at every point of $|z|=1$ for any function $f_n(z)$ meromorphic for $|z|<1$. A new sequence

$$\mathcal{J}_1(z), \mathcal{J}_2(z), \cdots \tag{0.3}$$

is thus determined. What are the relations between the sequences (0.1) and (0.3)?

The following questions are closely related to these two. Let $f(z)$ be a bounded function, analytic for $|z|<1$, with Fatou boundary function $F(z)$, as defined above. Let P be a point on $|z|=1$. Then what are the relations between $f(z)$ and $F(z)$ in a neighborhood of P? More generally let $f(z)$ be meromorphic for $|z|<1$. In §2 below, a boundary function $\mathcal{J}(z)$ of $f(z)$ will be defined at every point of $|z|<1$. What are the relations between $f(z)$ and $\mathcal{J}(z)$ in a neighborhood of P?

The purpose of this paper is to treat these four questions. Before treating them, however, a number of definitions, some new and some old, will be made in the following two sections.

1. Metric density and approximate continuity

In a previous paper‡ applications of the concepts of mean metric density and approximate continuity to complex function theory were made by the

* Presented to the Society, March 25 and March 26, 1932; received by the editors, February 6, 1932, and, in revised form, September 30, 1932.

† P. Fatou, Acta Mathematica, vol. 30 (1906), pp. 366–367.

‡ These Transactions, vol. 34 (1932), pp. 153–170.

author. The following lemma will be used in discussing further applications.

LEMMA 1.1. *Let E be a point set on the interval $-1<x<1$ having lower and upper mean metric density δ_l, δ_u respectively at $x=0$. Let E become E' under the transformation $x'=\Psi(x)$ and let E' have lower and upper mean metric density δ_l', δ_u' respectively at $\Psi(0)=0$.*

(a) *If $\Psi'(x)=d\Psi(x)/dx$ is continuous for $-1<x<1$, $\Psi'(0)>0$, then*

$$\delta_l' = \delta_l, \quad \delta_u' = \delta_u. \tag{1.11}$$

(b) *If $\Psi(x)=x^\nu$ for $x\geqq 0$, $\nu\geqq 1$, and $\Psi(x)=-|x|^\nu$ for $x\leqq 0$, then $\delta_u=1$ implies that $\delta_u'=1$.*

The simple proof of this lemma will be omitted here.*

COROLLARY. *If E has lower and upper metric densities on the right δ_{lr}, δ_{ur}, respectively at $x=0$ and if E' has lower and upper metric densities on the right δ_{lr}', δ_{ur}' respectively at $x=0$,*

$$\delta_{lr}' = \delta_{lr}, \quad \delta_{ur}' = \delta_{ur}, \tag{1.12}$$

in Case (a) *and $\delta_{ur}=1$ implies $\delta_{ur}'=1$ in Case* (b).

We can suppose that E has no points to the left of the origin, when the corollary follows immediately from the lemma.

Let $F(z)$ be a measurable function defined almost everywhere on $|z|=1$. The idea of approximate continuity will be slightly extended as follows. If the set of those points at which $|F(z)-\alpha|\leqq\epsilon$ has upper mean metric density 1 at z_0 for some complex number α and for all positive numbers ϵ, $F(z)$ will be said to be quasi-approximately continuous at z_0 with limit value α there. $F(z)$ may be quasi-approximately continuous at a point with several limit values there.

Let $F_1(z)$, $F_2(z)$, $\cdots$ be a sequence of measurable functions defined on a set of positive measure E on $|z|=1$. The sequence is said to converge in measure to a (measurable) function $F(z)$ when the measure of the set of those points for which $|F(z)-F_n(z)|\geqq\epsilon$ approaches 0 with $1/n$ for every positive number ϵ. If the sequence is uniformly bounded, one necessary and sufficient condition for this is that $F(z)$ be bounded and measurable on E and that

$$\lim_{n\to\infty}\int_E |F(z) - F_n(z)|\,|dz| = 0,$$

and another that every subsequence of the sequence $\{F_n(z)\}$ contain a further subsequence converging almost everywhere on E to $F(z)$.†

* Cf. the proof of Lemma 2.1 in the previous paper.

† F. Riesz, Paris Comptes Rendus, vol. 148 (1909), pp. 1303–1305.

LEMMA 1.2. *Let* $\{F_n(z)\}$ *be a sequence of measurable functions defined on a measurable set E on* $|z|=1$, $mE>0$. *A necessary and sufficient condition that the sequence converge in measure on E to the measurable function* $F(z)$ *is that*

$$\lim_{n\to\infty} \underline{B}\{\,|F(z)-F_n(z)|\,, E_n\} = 0^*$$

for every sequence $\{E_n\}$ *of measurable point sets on* $|z|=1$ *such that* $E_n\subset E$, $n=1, 2, \cdots$, *and such that*

$$\liminf_{n\to\infty} mE_n > 0.$$

This result is an immediate consequence of the definition of convergence in measure.

It will be seen in §5 that the concepts of approximate continuity and convergence in measure are related to each other.

2. CLUSTER VALUES OF FUNCTIONS AND OF SEQUENCES

In the following, points of the extended plane, or of the sphere corresponding to it by stereographic projection, will be considered. "Closed," "open," etc., used of point sets of the plane, will refer to the corresponding point sets on the sphere. The point ∞ is then in no way exceptional, and is allowable as a value assumed by a function.

Let $f(z)$ be a single-valued function defined in a domain† γ bounded by a simple closed Jordan curve Γ (i.e., a one-to-one and continuous image of the perimeter of a circle). Let P be a point on Γ. Then if there is a complex number α and a sequence of points $\{z_n\}$, in γ, such that

$$\lim_{n\to\infty} z_n = P, \quad \lim_{n\to\infty} f(z_n) = \alpha, \tag{2.01}$$

α is called a cluster value of $f(z)$ in γ at P. The set of all cluster values of $f(z)$ in γ at P is called the cluster set of $f(z)$ in γ at P. This set is closed and connected if $f(z)$ is continuous in γ. The function $\mathcal{J}(z)$, defined for every point P on Γ, as the cluster set of $f(z)$ in γ at P will be called the cluster boundary function of $f(z)$. It is evidently multiple-valued, in general. The function $f(z)$ is said to have the cluster value α on a given path to P if there exists a sequence of points $\{z_n\}$ on that path, so that (2.01) is satisfied. If γ is the interior of the unit circle, $|z|<1$, the path will be called non-tangential if it is contained in some angle with vertex at P whose sides are chords of $|z|=1$.

---

* Throughout this paper if $F(z)$ is a function defined on a set E, $\underline{B}\{\,|F(z)|, E\}$ will denote the greatest lower bound of $|F(z)|$ on E, and $O\{F(z), E\}$ will denote the oscillation of $F(z)$ on E, i.e. the least upper bound of $|F(P)-F(Q)|$ for P, Q any two points of E.

† In this paper, any open connected point set will be called a domain.

If the path is a continuous curve C* and if there is only a single cluster value of $f(z)$ on C:

$$\lim_{z\to P} f(z) = \alpha \tag{2.02}$$

when z approaches P on C, $f(z)$ is said to have the convergence value α at P.

If there is a complex number α, a sequence of points $\{z_n\}$ on Γ in a neighborhood of P, all on one side of P and different from P, such that

$$\lim_{n\to\infty} z_n = P, \quad \lim_{n\to\infty} \mathcal{J}(z_n) = \alpha \tag{2.03}$$

(choosing one definite value of $\mathcal{J}(z_n)$ for each value of n), α is called a cluster value of $f(z)$ on Γ at P on the side in question. The cluster sets of $f(z)$ on Γ at P on each side are then defined as the set of all the cluster values of $f(z)$ on that side, and the cluster set of $f(z)$ on Γ at P is the sum of these two sets. If $f(z)$ is continuous in γ, the cluster sets of $f(z)$ on Γ at P on each side are closed and connected. If E is a point set on Γ which has P as a limit point, and if in (2.03) the points $\{z_n\}$ all belong to E, the set of all values α thus determined will be called the cluster set of $f(z)$ on Γ on E at P. These ideas were introduced by Painlevé.†

It does not seem to have been realized that the above definitions are analogues of certain definitions for sequences of functions, defined in the interior of the unit circle. Let

$$f_1(z), f_2(z), \cdots \tag{2.04}$$

be a sequence of single-valued functions defined for $|z|<1$, with cluster boundary functions $\mathcal{J}_1(z)$, $\mathcal{J}_2(z)$, $\cdots$ respectively on $|z|=1$. Then if there is a complex number α, a subsequence $\{f_{a_n}(z)\}$, and a sequence of points $\{z_{a_n}\}$ in $|z|<1$, such that

$$\lim_{n\to\infty} f_{a_n}(z_{a_n}) = \alpha, \tag{2.05}$$

α will be called a cluster value of the sequence (2.04) in $|z|<1$. If $g_n(z)$ is defined by

$$g_n(z) = f_{a_n}\left(\frac{z - z_{a_n}}{\bar{z}_{a_n} z - 1}\right)^{\ddagger} \tag{2.06}$$

and if

* That is, C is determined by $z=\psi(t)$ where $\psi(t)$ is continuous for $0\leq t\leq 1$, $z=\psi(t)$ is in γ for $0\leq t<1$, $\psi(1)=P$.

† P. Painlevé, Paris Comptes Rendus, vol. 131 (1900), p. 489.

‡ The conjugate complex number of ξ is denoted by $\bar{\xi}$.

$$\lim_{n\to\infty} g_n(z) = \alpha \tag{2.07}$$

uniformly in every closed subregion of $|z| <1$, α will be called a convergence value of the sequence. The sets of all cluster and convergence values of the sequence in $|z| <1$ will be called the cluster and convergence sets in $|z| <1$, respectively. The former set is closed.

If there is a complex number α, a subsequence $\{f_{a_n}(z)\}$, and a sequence of points $\{z_{a_n}\}$ on $|z| =1$, such that

$$\lim_{n\to\infty} \mathcal{J}_{a_n}(z_{a_n}) = \alpha \tag{2.08}$$

(choosing one definite value for $\mathcal{J}_{a_n}(z_{a_n})$ for each value of n), α will be called a cluster value of the sequence (2.04) on $|z| =1$. The set of all cluster values of the sequence on $|z| =1$, which we designate as the cluster set of the sequence on $|z| =1$, is closed.

Let $\{A_n\}$ be a set of open arcs on $|z| =1$. Then if α is a cluster value of the sequence (2.04) in $|z| <1$ in accordance with the definition given above and if under the transformation

$$z' = \frac{z - z_{a_n}}{\bar{z}_{a_n} z - 1}, \tag{2.09}$$

the arc A_{a_n} becomes the arc A_{a_n}' such that

$$\liminf_{n\to\infty} mA_{a_n}' > 0, \tag{2.10}$$

α will be called a cluster value of the sequence in $|z| <1$ with respect to the arcs $\{A_n\}$. If in particular

$$\lim_{n\to\infty} mA_{a_n}' = 2\pi, \tag{2.11}$$

α will be called a strong cluster value of the sequence in $|z| <1$ with respect to the arcs $\{A_n\}$. If

$$\limsup_{n\to\infty} | z_{a_n} | < 1,$$

the conditions (2.10) and (2.11) are equivalent to

$$\liminf_{n\to\infty} mA_{a_n} > 0, \quad \lim_{n\to\infty} mA_{a_n} = 2\pi,$$

respectively. If

$$A_n = A_1,\ n > 1,\ mA_1 < 2\pi, \tag{2.12}$$

then in the case (2.10), a convergent subsequence of the sequence $\{z_{a_n}\}$ must

approach (i) a point of $|z|<1$, or (ii) a point of A_1, or (iii) an end point of A_1, remaining in the angle between A_1 and some chord through that end point. In the case (2.11), still assuming (2.12), a convergent subsequence of the sequence $\{z_{a_n}\}$ must approach (i) a point of A_1, or (ii) an end point of A_1, approaching the end point tangentially—on the same side of the end point as A_1. Conversely the conditions given are sufficient that α be a cluster value or a strong cluster value of the sequence $\{f_n(z)\}$ with respect to the arcs $\{A_n\}$ respectively. There is only slight modification of these criteria if $\liminf_{n\to\infty} mA_n>0$. The sets of all cluster values and strong cluster values with respect to a set of arcs will be called the cluster set and the strong cluster set of the sequence with respect to those arcs, respectively. It is not hard to show that the latter is a closed subset of the former. If α is a cluster value with respect to a set of arcs on $|z|=1$, i.e. if (2.05) and (2.10) are satisfied, and if (2.07) is also satisfied, α will be called a convergence value of the sequence with respect to those arcs. The convergence set with respect to the arcs will then be the set of all these convergence values. The convergence set with respect to a set of arcs is a subset of the strong cluster set with respect to the arcs. For if α is a convergence value with respect to the arcs $\{A_n\}$ we can suppose that $\liminf_{n\to\infty} mA_{a_n}>0$ (or we could use the sequence determined by (2.06)). By (2.07) there is then a subsequence $\{f_{b_n}(z)\}$ of $\{f_{a_n}(z)\}$ and a sequence of points $\{\xi_{b_n}\}$ such that $\lim_{n\to\infty}|\xi_{b_n}|=1$ and such that $\lim_{n\to\infty} f_{b_n}(\xi_{b_n})=\alpha$. We can suppose that ξ_{b_n} is so chosen that the distance from ξ_{b_n} to the midpoint of the arc A_{b_n} approaches 0 with $1/n$. Then α is a strong cluster value of the sequence with respect to the arcs $\{A_n\}$, by the criterion suggested above.

If α is a cluster value of the sequence (2.04) on $|z|=1$ in accordance with (2.08) and if the point z_{a_n} lies on the arc A_{a_n} for all values of n, α will be called a cluster value of the sequence on $|z|=1$ on the arcs $\{A_n\}$. The set of all these cluster values will be called the cluster set of the sequence on the arcs considered. This set is closed.

A point α will be said to be assumed by the sequence (2.04) if every function of the sequence except for at most a finite number assumes the value α. A point α will be said to be exceptional to or omitted by the sequence if at most a finite number of the functions assume the value α.

3. The properties of the boundary functions of a uniformly bounded convergent sequence of analytic functions

Let

$$f_1(z),\ f_2(z),\ \cdots \tag{3.01}$$

be a uniformly bounded sequence of functions analytic for $|z|<1$, with Fatou boundary functions

$$F_1(z), F_2(z), \cdots \tag{3.02}$$

respectively. What are the relations between these two sequences? The sequence (3.01) forms a normal family.* If it is uniformly convergent in every closed subregion of $|z|<1$ to the limit function $f(z)$, we can reduce the problem to that in which the limit function vanishes identically by substituting the sequence $\{f_n(z)-f(z)\}$ for $\{f_n(z)\}$. This will be convenient in much of what follows. The problem solved in this section is the following. Necessary and sufficient conditions are found on the sequence (3.02) that the sequence (3.01) converge uniformly in every closed subregion of $|z|<1$ to a function $f(z)$, where, if convenient, the limit function $f(z)$ is supposed to vanish identically. It will be seen that any domain bounded by a simple closed rectifiable Jordan curve could be used instead of $|z|<1$ as the domain of definition of the functions of the sequence (3.01).

LEMMA 3.1. *Let $f_1(z), f_2(z), \cdots$ be a uniformly bounded sequence of functions analytic for $|z|<1$, with Fatou boundary functions $F_1(z)$, $F_2(z), \cdots$ respectively, $|F_n(z)| \leqq 1$, $n=1, 2, \cdots$. Then if there is a sequence of points $\{z_n\}$ such that*

$$\limsup_{n\to\infty} |z_n| < 1, \quad \lim_{n\to\infty} |f_n(z_n)| = 1, \tag{3.101}$$

it follows that

$$\lim_{n\to\infty} |f_n(z)| = 1 \tag{3.102}$$

uniformly in every closed subregion of $|z|<1$ and that the sequence $\{|F_n(z)|\}$ converges in measure to 1 *on $|z|=1$.*

The sequence $\{f_n(z)\}$ forms a normal family, as remarked above, and any limit function $f(z)$ must satisfy the two inequalities

$$|f(z)| \leqq 1, \quad |f(z_0)| = 1, \quad |z_0| < 1, \tag{3.103}$$

where z_0 is a limit point of the sequence $\{z_n\}$. It follows from the maximum principle that $|f(z)| \equiv 1$. Then every limit function of the sequence is a constant of modulus 1. If the sequence $\{|f_n(z)|\}$ did not converge uniformly to 1 in every closed subregion of $|z|<1$, there would be a closed subregion R, and a positive number $\rho<1$, such that

$$|f_{a_n}(\xi_{a_n})| \leqq \rho \qquad (n = 1, 2, \cdots), \tag{3.104}$$

* P. Montel, *Leçons sur les Familles Normales*, Paris, 1927, p. 21.

for some subsequence $\{f_{a_n}(z)\}$ and a sequence $\{\xi_{a_n}\}$ of points in R. From $\{f_{a_n}(z)\}$ can be extracted a further subsequence $\{f_{b_n}(z)\}$ such that $\lim_{n\to\infty} |f_{b_n}(z)|=1$ uniformly in R, contradicting (3.104). The sequence $\{|f_n(z)|\}$ therefore converges uniformly to 1 in every closed subregion of $|z|<1$. In particular $\lim_{n\to\infty} |f_n(0)|=1$. Now by the Cauchy integral formula,

$$(3.105) \qquad | f_n(0) | \leqq \frac{1}{2\pi}\int_{|z|=1} | F_n(z) | | dz | \leqq 1,$$

so that

$$(3.106)\ 0 = \lim_{n\to\infty} [1 - | f_n(0) |] \geqq \lim_{n\to\infty} \frac{1}{2\pi}\int_{|z|=1} [1 - | F_n(z) |] | dz | = 0,$$

which proves that the sequence $\{|F_n(z)|\}$ converges in measure to 1 on $|z|=1$.*

THEOREM 3.1. *Let $f_1(z), f_2(z), \cdots$ be a uniformly bounded sequence of functions analytic for $|z|<1$, with Fatou boundary functions $F_1(z)$, $F_2(z), \cdots$ respectively, $|F_n(z)|\leqq 1$, $n=1, 2, \cdots$.*

(a) *If there is a sequence of points $\{z_n\}$ such that*

$$(3.111) \qquad \limsup_{n\to\infty} | z_n | < 1, \quad \lim_{n\to\infty} f_n(z_n) = \alpha, \ | \alpha | = 1,$$

it follows that

$$(3.112) \qquad \lim_{n\to\infty} f_n(z) = \alpha$$

uniformly in every closed subregion of $|z|<1$ and that the sequence $\{F_n(z)\}$ converges in measure to α on $|z|=1$.†

(b) *If the sequence $\{F_n(z)\}$ converges in measure to α, $|\alpha|\leqq 1$, on a measurable set E on $|z|=1$, $mE>0$, the conclusions of* (a) *hold.*

If we take $|\alpha|=1$ in (b), the condition of (b) is necessary and sufficient that (3.112) hold, so Theorem 3.1 solves the given problem in a very particular case.

(a) We can assume that $\alpha=1$. The result (a) is then simply Lemma 3.1 applied to the sequence $\{\phi_n(z)\}$, where

$$(3.113) \qquad \phi_n(z) = e^{f_n(z)-1}.$$

* Cf. §1.

† A similar theorem was proved by J. L. Walsh (who uses the term quasi-convergence instead of convergence in measure), and applied in another connection, these Transactions, vol. 32 (1930), pp. 378–379.

(b) To prove (b) it is only necessary to prove that the hypothesis of (b) implies (3.112). By a theorem of Khintchine and Ostrowski* the actual convergence of the sequence $\{F_n(z)\}$ almost everywhere on E implies (3.112). The proof as given in Bieberbach's book also proves the more general result desired. A still more general result will be useful, however. *Let $E_n(\epsilon)$ be the set of those points on $|z|=1$ for which $|F_n(z)-\alpha|\leqq\epsilon$. Then it is sufficient in* (b) *if*

$$\lim_{\epsilon\to 0}\left\{\limsup_{n\to\infty}\epsilon^{mE_n(\epsilon)}\right\}=0. \tag{3.114}$$

This follows from the Ostrowski-Nevanlinna inequality, or the proof referred to above can be modified to prove this also (by choosing the constant A used in it properly). It is sufficient for (3.114) that

$$\liminf_{\epsilon\to 0}\left[\liminf_{n\to\infty} mE_n(\epsilon)\right]>0,$$

and it is this special case which will be used most in the applications in this paper.

Corollary. *In the above theorem if $w_n=f_n(z_n)$ for large values of n: $n\geqq n(\rho)$, is outside every circle C_ρ tangent to $|w|=1$ at $w=\alpha$, of radius $\rho<1$, the same will be true of the values of $w=f_n(z)$ for z in any fixed closed subregion of $|z|<1$ and the measure of the set of those points on $|z|=1$ at which $w=F_n(z)$ is inside C_ρ approaches* 0 *with $1/n$ for every value of $\rho<1$.*

We can suppose that $\alpha=1$. The corollary is simply the theorem applied to the sequence $\{\phi[f_n(z)]\}$ where $\phi(w)$ is defined by

$$\phi(w)=e^{(w+1)/(w-1)}.$$

Theorem 3.2. *Let $f_1(z), f_2(z), \cdots$ be a uniformly bounded sequence of functions analytic for $|z|<1$, with Fatou boundary functions $F_1(z), F_2(z), \cdots$ respectively. Suppose that $f_n(z)\neq 0$, $n=1, 2, \cdots$.*

(a) *If there is a sequence of points $\{z_n\}$ such that*

$$\limsup_{n\to\infty}|z_n|<1,\quad \lim_{n\to\infty}f_n(z_n)=0, \tag{3.21}$$

it follows that

$$\lim_{n\to\infty}f_n(z)=0 \tag{3.22}$$

* See, for example, L. Bieberbach, *Lehrbuch der Funktionentheorie*, 2d edition, vol. II, 1931, pp. 157–158.

uniformly in every closed subregion of $|z|<1$ *and that the sequence* $\{1/\log F_n(z)\}$* *converges in measure to* 0 *on* $|z|=1$ *whatever branch of* $\log F_n(z)$ *is chosen.*

(b) *If the sequence* $\{1/\log F_n(z)\}$ *converges in measure to* 0 *on a measurable set E of positive measure on* $|z|=1$, *it follows that*

$$\lim_{n\to\infty} 1/\log f_n(z) = 0 \tag{3.23}$$

uniformly in every closed subregion of $|z|<1$ *and that the sequence* $\{1/\log F_n(z)\}$ *is convergent in measure to* 0 *on* $|z|=1$, *where the branch of* $\log F_n(z)$ *is determined by that of* $\log f_n(z)$ *in* (3.23).

The uniformly bounded sequence $\{f_n(z)\}$ forms a normal family. Let $f(z)$ be a limit function of the family: $f(z)=\lim_{n\to\infty} f_{a_n}(z)$. Then $f(z_0)=0$ in Case (a), where z_0 is a limit point of the sequence $\{z_n\}$. Then by a well known theorem of Hurwitz, if $f(z)\not\equiv 0$, $f_{a_n}(z)$ must vanish in a neighborhood of z_0 for all large values of n. Since this is not true, $f(z)\equiv 0$. Thus every limit function of the family vanishes identically, and (3.22) is proved by an argument similar to that used in proving (3.102) in the proof of Lemma 3.1.

We can suppose that $|f_n(z)|<1$, $n=1, 2, \cdots$. Define the function $\phi_n(z)$, analytic for $|z|<1$, by

$$\phi_n(z) = \frac{\log f_n(z) + 1}{\log f_n(z) - 1}\cdot \tag{3.24}$$

Then $|\phi_n(z)|<1$ and $\phi_n(z)$ has the boundary function $\Phi_n(z)$:

$$\Phi_n(z) = \frac{\log F_n(z) + 1}{\log F_n(z) - 1}\cdot \tag{3.25}$$

Now $\lim_{n\to\infty} \phi_n(z)=1$ and $\lim_{n\to\infty} 1/\log f_n(z)=0$ are equivalent statements, so the theorem is an immediate consequence of Theorem 3.1. We note a generalization of (b) corresponding to one of Theorem 3.1 (b) which will be used in proving the next theorem. *Let* $E_n(\epsilon)$ *be the set of those points on* $|z|=1$ *for which* $|\log F_n(z)|>1/\epsilon$. *It is sufficient in* (b) *if*

$$\liminf_{\epsilon\to 0}\left[\liminf_{n\to\infty} mE_n(\epsilon)\right] > 0.$$

* Choose a branch of $\log f_n(z)$ at some point of $|z|<1$ and continue it analytically throughout $|z|<1$, determining a single-valued analytic function which has finite radial boundary values wherever $F_n(z)\neq 0$. Since $f_n(z)\not\equiv 0$, $F_n(z)=0$ at most on a set of measure 0 by a theorem of F. and M. Riesz. Then this branch of $\log f_n(z)$ has a finite-valued and single-valued boundary function defined almost everywhere on $|z|=1$ which will be denoted by $\log F_n(z)$. The function $\log F_n(z)$ has infinitely many branches differing by integral multiples of 2π.

The condition that the sequence $\{1/\log F_n(z)\}$ converge in measure to 0 on $|z|=1$ is equivalent to

$$\lim_{n\to\infty} \underline{B}\{\,|\,1/\log F_n(z)\,|\,,\,E_n\} = 0, \tag{3.26}$$

for every sequence $\{E_n\}$ of measurable sets on $|z|=1$ such that

$$\liminf_{n\to\infty} mE_n > 0,$$

by Lemma 1.2. In this form the result can be more readily compared with that in the next theorem.

An example of the theorem in which $|F_n(z)|=1$ except at $z=1$, $F_n(1)=0$, for all values of n is given by

$$f_n(z) = e^{n(z+1)/(z-1)} \qquad (n = 1, 2, \cdots). \tag{3.27}$$

Theorem 3.2 enables us to solve a particular case of the problem proposed at the beginning of this section which includes the particular case of Theorem 3.1 obtained by setting $|\alpha|=1$ in (b). It will be seen that this particular case has important applications.

THEOREM 3.3. *Let $f_1(z), f_2(z), \cdots$ be a uniformly bounded sequence of functions analytic for $|z|<1$ with Fatou boundary functions $F_1(z), F_2(z), \cdots$ respectively. Suppose that $f_n(z)\neq 0$, $n=1, 2, \cdots$.*

(a) *If there is a sequence of points $\{z_n\}$ such that*

$$\limsup_{n\to\infty} |\,z_n\,| < 1, \quad \lim_{n\to\infty} f_n(z_n) = 0, \tag{3.301}$$

it follows that $\lim_{n\to\infty} f_n(z)=0$ uniformly in every closed subregion of $|z|<1$ and that

$$\lim_{n\to\infty} \frac{\underline{B}\{\,|\,F_n(z)\,|\,,\,E_n\}}{1 + O\{\operatorname{arc} F_n(z),\,E_n\}} = 0^* \tag{3.302}$$

for every sequence $\{E_n\}$ of measurable point sets on $|z|=1$ satisfying

$$\liminf_{n\to\infty} mE_n > 0. \tag{3.303}$$

(b) *If there is a sequence of measurable point sets $\{\mathcal{E}_n\}$ on $|z|=1$ satisfying* (3.303) *and such that every sequence $\{E_n\}$ of measurable point sets on $|z|=1$ such that $E_n\subset\mathcal{E}_n$, $n=1, 2, \cdots$, which satisfies* (3.303) *also satisfies* (3.302), *then $\lim_{n\to\infty} f_n(z)=0$ uniformly in every closed subregion of $|z|<1$.*

---

* Cf. the first note on p. 420. Arc $F_n(z)$ can be defined as the imaginary part of $\log F_n(z)$: $\mathfrak{I}\log F_n(z)$. Its oscillation on E_n is independent of the branch of $\log F_n(z)$ chosen.

The statement (b) is stronger than the converse of (a). This theorem shows what happens if $\lim_{n\to\infty} f_n(z)=0$ under the above circumstances and if the sequence of boundary functions does not converge in measure to 0. The condition (3.302) is only slightly stronger than (3.26) as is to be expected.

(a) Suppose that (3.301) is satisfied. By the previous theorem

$$\lim_{n\to\infty} f_n(z) = 0$$

uniformly in every closed subregion of $|z|<1$. Unless (3.302) is true, there is a subsequence $\{F_{a_n}(z)\}$, a positive number λ and a sequence $\{E_{a_n}\}$ of measurable point sets on $|z|=1$ such that $\liminf_{n\to\infty} mE_{a_n}>0$ and such that

$$\frac{\underline{B}\{\,|F_{a_n}(z)|\,,\,E_{a_n}\}}{1+O\{\operatorname{arc} F_{a_n}(z),\,E_{a_n}\}} \geqq \lambda \qquad (n = 1, 2, \cdots). \tag{3.304}$$

Let P_{a_n} be a point of E_{a_n} and choose $\log F_{a_n}(z)$ so that

$$|\,\mathfrak{J}\log F_{a_n}(P_{a_n})\,| \leqq \pi \qquad (n = 1, 2, \cdots). \tag{3.305}$$

Then, since $O\{\operatorname{arc} F_{a_n}(z),\, E_{a_n}\} \leqq M/\lambda$ by (3.304), where $|F_n(z)| \leqq M$, $n=1, 2, \cdots$,

$$|\,\mathfrak{J}\log F_{a_n}(z)\,| \leqq \pi + M/\lambda \qquad (n = 1, 2, \cdots), \tag{3.306}$$

on E_{a_n}. Now by (3.304), $|F_{a_n}(z)| \geqq \lambda$ on E_{a_n}. Then

$$|\log F_{a_n}(z)\,| \leqq |\log\lambda| + |\log M| + \pi + M/\lambda \qquad (n = 1, 2, \cdots), \tag{3.307}$$

on E_{a_n}. But by Theorem 3.2 the sequence $\{1/\log F_{a_n}(z)\}$ converges in measure to 0 on $|z|=1$, so this inequality is impossible.

(b) Suppose that the hypotheses of (b) are satisfied. Determine $\log F(z)$ from a branch of $\log f_n(z)$ for which

$$|\,\mathfrak{J}\log f_n(0)\,| \leqq \pi \qquad (n = 1, 2, \cdots). \tag{3.308}$$

Then it is sufficient to show that the measure of the subset of $\mathcal{E}_n$ on which $|\log F_n(z)| \leqq K$ approaches 0 with $1/n$ for every value of K. For then, by Theorem 3.2 (b) in its generalized form, $\lim_{n\to\infty} 1/\log f_n(0)=0$, which implies, by (3.308), that $\lim_{n\to\infty} f_n(0)=0$. This is sufficient that $\lim_{n\to\infty} f_n(z)=0$ uniformly in every closed subregion of $|z|<1$, by Theorem 3.2 (a). Thus if (b) were not true there would be a number K, and a subsequence $\{F_{a_n}(z)\}$, such that

$$|\log F_{a_n}(z)\,| \leqq K \qquad (n = 1, 2, \cdots), \tag{3.309}$$

on a subset E_{a_n} of $\mathcal{E}_{a_n}$, where $\liminf_{n\to\infty} mE_{a_n}>0$. Then

(3.310) $$|F_{a_n}(z)| \geqq e^{-K}$$

on E_{a_n}, i.e.

$$\underline{B}\{|F_{a_n}(z)|, E_{a_n}\} \geqq e^{-K}$$

so, by (3.302),

(3.311) $$\lim_{n\to\infty} O\{\operatorname{arc} F_{a_n}(z), E_{a_n}\} = +\infty.$$

Then there are points P_{a_n}, Q_{a_n} on E_{a_n} for large values of n such that

(3.312) $$|\operatorname{arc} F_{a_n}(P_{a_n}) - \operatorname{arc} F_{a_n}(Q_{a_n})| \geqq 4K.$$

But then either $|\log F_{a_n}(P_{a_n})| \geqq 2K$ or $|\log F_{a_n}(Q_{a_n})| \geqq 2K$, which contradicts (3.309). The theorem is thus completely proved.

Now consider the general problem proposed at the beginning of the section. Necessary and sufficient conditions are to be found on the sequence (3.02) that the sequence (3.01) converge uniformly to 0 in every closed subregion of $|z|<1$. The problem has just been solved if $f_n(z)\neq 0$ except for a finite number of values of n. The following theorem gives the general solution.

THEOREM 3.4. *Let $f_1(z), f_2(z), \cdots$ be a uniformly bounded sequence of functions analytic for $|z|<1$, with Fatou boundary functions $F_1(z)$, $F_2(z), \cdots$ respectively. Let $\{A_n\}$, $\{A_n'\}$ be sequences of arcs on $|z|=1$ such that A_n, A_n' have no points in common and such that*

(3.401) $$\liminf_{n\to\infty} mA_n > 0, \quad \liminf_{n\to\infty} mA_n' > 0.$$

Then there exist two sequences of uniformly bounded functions $g_1(z), g_2(z), \cdots$, $h_1(z)$, $h_2(z), \cdots$ analytic for $|z|<1$ with Fatou boundary functions $G_1(z)$, $G_2(z), \cdots, H_1(z), H_2(z), \cdots$ respectively such that

(3.402) $$f_n(z) = g_n(z)\cdot h_n(z) \qquad (n = 1, 2, \cdots),$$

and such that $g_n(z)\neq 0$ in $\mathfrak{S}(A_n)$, $h_n(z)\neq 0$ in $\mathfrak{S}(A_n')$. A necessary and sufficient condition that*

(3.403) $$\lim_{n\to\infty} f_n(z) = 0$$

uniformly in every closed subregion of $|z|<1$ is that

---

* It will be convenient in this and the following sections if A is an arc of the unit circle, to denote by $\mathfrak{S}(A)$ the interior of the segment bounded by A and by its chord.

$$\lim_{n\to\infty} \frac{\underline{B}\{\,|G_n(z)|\,, E_n\}}{1+O\{\operatorname{arc} G_n(z), E_n\}}\cdot\frac{\underline{B}\{\,|H_n(z)|\,, E_n'\}}{1+O\{\operatorname{arc} H_n(z), E_n'\}} = 0^* \tag{3.404}$$

for every pair of sequences $\{E_n\}$, $\{E_n'\}$ *of measurable point sets on* $|z|=1$ *such that* $E_n \subset A_n$, $E_n' \subset A_n'$, $n=1, 2, \cdots$, *and such that*

$$\liminf_{n\to\infty} mE_n > 0, \quad \liminf_{n\to\infty} mE_n' > 0. \tag{3.405}$$

We consider only the case in which A_n and A_n' are the same arcs A and A' respectively for all values of n. The general case can be proved by a slight modification of the following proof.

Let $\phi_1(w)$, $\phi_2(w)$ be functions mapping $|w|<1$ in a one-to-one and conformal way on $\mathfrak{S}(A)$ and $\mathfrak{S}(A')$ respectively. These functions can easily be determined explicitly. Let

$$\alpha_1^{(n)}, \alpha_2^{(n)}, \cdots, |\alpha_1^{(n)}| \leqq |\alpha_2^{(n)}| \leqq \cdots, \tag{3.406}$$

be the zeros of $f_n(z)$ in the interior of $\mathfrak{S}(A)$, where the non-simple zeros appear in the list a number of times equal to their multiplicity. By a theorem of Blaschke†

$$h_n(z) = \prod_{j=1}^{\infty} \frac{z-\alpha_j^{(n)}}{\bar{\alpha}_j^{(n)} z - 1}\bar{\alpha}_j^{(n)} \tag{3.407}$$

defines a bounded function, analytic for $|z|<1$, where the product converges uniformly in every closed subregion of $|z|<1$. Define the function $g_n(z)$ by

$$g_n(z) = f_n(z)/h_n(z). \tag{3.408}$$

Then it is readily seen that $g_n(z)$ is a bounded function, analytic in the interior of the unit circle. The zeros of the functions $g_n(z)$, $h_n(z)$ have the required properties.

Equation (3.404) is equivalent to the following:

$$\lim_{n\to\infty} g_n[\phi_1(w)]\cdot h_n[\phi_2(w)] = 0, \tag{3.409}$$

uniformly in every closed subregion of $|w|<1$. For suppose that (3.409) is true. If (3.404) is not also true there are subsequences $\{g_{a_n}(z)\}$, $\{h_{a_n}(z)\}$ (which we can suppose convergent in $|z|<1$ since the sequences $\{g_n(z)\}$,

---

* Each branch of arc $g_n(z)$ is a single-valued function in $\mathfrak{S}(A_n)$, thus determining a single-valued branch of arc $G_n(z)$. There are single-valued branches of arc $H_n(z)$ on A_n' by the same argument.

† See, for example, Montel, *Leçons sur les Familles Normales*, Paris, 1927, p. 180. If $\mathfrak{S}(A)$ includes $z=0$ and if $f_n(z)$ has a zero of order λ_n there, the product is taken from $j=\lambda_n+1$ to ∞ and the factor z^{λ_n} replaces the first λ_n factors.

$\{h_n(z)\}$ are normal families) and sets $\{E_{a_n}\}$, $\{E_{a_n}'\}$ satisfying (3.405) such that

$$(3.410)\qquad \liminf_{n\to\infty} \frac{\underline{B}\{\,|G_{a_n}(z)|\,,E_{a_n}\}}{1+O\{\operatorname{arc} G_n(z),\,E_{a_n}\}}\cdot\frac{\underline{B}\{\,|H_{a_n}(z)|\,,E_{a_n}'\}}{1+O\{\operatorname{arc} H_{a_n}(z),\,E_{a_n}'\}} > 0.$$

Let the point sets on $|w|=1$ transformed into E_{a_n} and E_{a_n}' on $|z|=1$ by the transformations $z=\phi_1(w)$, $z=\phi_2(w)$ be $\mathcal{E}_{a_n}$, $\mathcal{E}_{a_n}'$ respectively. Then it is easily seen that

$$(3.411)\qquad \liminf_{n\to\infty} m\,\mathcal{E}_{a_n} > 0,\quad \liminf_{n\to\infty} m\,\mathcal{E}_{a_n}' > 0.$$

Inequality (3.410) is the same as

$$(3.412)\ \liminf_{n\to\infty} \frac{\underline{B}\{\,|G_{a_n}[\phi_1(w)]|\,,\mathcal{E}_{a_n}\}}{1+O\{\operatorname{arc} G_{a_n}[\phi_1(w)],\,\mathcal{E}_{a_n}\}}\cdot\frac{\underline{B}\{\,|H_{a_n}[\phi_2(w)]|\,,\mathcal{E}_{a_n}'\}}{1+O\{\operatorname{arc} H_{a_n}[\phi_2(w)],\,\mathcal{E}_{a_n}'\}} > 0.$$

But this means, by Theorem 3.3, that neither of the (convergent) sequences $\{g_{a_n}[\phi_1(w)]\}$, $\{h_{a_n}[\phi_2(w)]\}$ can converge to the function 0, which contradicts (3.409). Conversely suppose that (3.404) is true. If (3.409) is not true, there are subsequences $\{g_{a_n}[\phi_1(w)]\}$, $\{h_{a_n}[\phi_2(w)]\}$ which are convergent to functions which do not vanish identically. Then by Theorem 3.3 there are sequences of sets $\{\mathcal{E}_{a_n}\}$, $\{\mathcal{E}_{a_n}'\}$ on $|w|=1$ such that (3.411) and (3.412) are satisfied. Letting the sets $\{E_{a_n}\}$, $\{E_{a_n}'\}$ on $|z|=1$ correspond as above to the sets $\{\mathcal{E}_{a_n}\}$, $\{\mathcal{E}_{a_n}'\}$, (3.404) is contradicted. Equations (3.409) and (3.404) are thus equivalent.

Equations (3.409) and (3.403) are also equivalent. For suppose that (3.403) is true. We have to show that every limit function of the sequence $\{g_n[\phi_1(w)]\cdot h_{a_n}[\phi_2(w)]\}$ is the function 0. Suppose that this is not the case. Let $\{g_{a_n}[\phi_1(w)]\cdot h_{a_n}[\phi_2(w)]\}$ be a convergent subsequence, not converging to the function 0. We can suppose further that the sequences $\{g_{a_n}[\phi_1(w)]\}$, $\{h_{a_n}[\phi_2(w)]\}$ are also convergent, say to $g(w)$, $h(w)$ respectively. Since $g(w)h(w)\not\equiv 0$, $g(w)\not\equiv 0$, and $h(w)\not\equiv 0$. Then the sequences $\{g_{a_n}(z)\}$, $\{h_{a_n}(z)\}$ converge in $\mathcal{S}(A)$, $\mathcal{S}(A')$ respectively to functions which do not vanish identically. By a theorem of Stieltjes* these sequences are convergent throughout $|z|<1$ (to functions which cannot vanish identically). Then the sequence $\{f_{a_n}(z)\}$ cannot converge to 0, contrary to the hypothesis that (3.403) is true. Conversely suppose that (3.409) is true. We must show that (3.403) is true, i.e. that the only limit function of the sequence $\{f_n(z)\}$ is the function 0. Suppose $\{f_{a_n}(z)\}$ were a subsequence of $\{f_n(z)\}$ not converging to

* See for example P. Montel, loc. cit., pp. 28–30.

0. We can suppose that the sequences $\{g_{a_n}(z)\}$, $\{h_{a_n}(z)\}$ are both convergent. They converge to functions not vanishing identically, so the sequences $\{g_{a_n}[\phi_1(w)]\}$, $\{h_{a_n}[\phi_2(w)]\}$ have the same property, contradicting (3.409).

It has thus been shown that (3.409), (3.404), and (3.409), (3.403), are pairs of equivalent statements, proving the theorem.

4. The properties of the cluster boundary functions of a sequence of meromorphic functions

Let

$$f_1(z), f_2(z), \cdots \tag{4.01}$$

be a sequence of functions meromorphic for $|z| < 1$, with cluster boundary functions

$$F_1(z), F_2(z), \cdots \tag{4.02}$$

respectively, as defined in §2. The problem to be attacked in this section is that of finding the relations between these two sequences.

Theorem 4.1. *Let $f_1(z), f_2(z), \cdots$ be a sequence of functions meromorphic for $|z| < 1$. Let the cluster sets of the sequence in $|z| < 1$ and on $|z| = 1$ be s and S respectively. Then if there is a point α belonging to s but not to S, no point of the domain D containing α and bounded only by points of S* is omitted by the sequence.*

This theorem is proved by an application of the maximum principle for analytic functions which is fairly obvious, so the proof will be omitted. The theorem is stated only to allow ready comparison with Theorem 4.2, the principal result of this section. To prove Theorem 4.2, which generalizes Theorem 4.1, we need a succession of lemmas.

Lemma 4.1. *Let $\{f_n(z)\}$ be a uniformly bounded sequence of functions analytic for $|z| < 1$. Let $\{A_n\}$ be a sequence of arcs on $|z| = 1$,*

$$\liminf_{n\to\infty} mA_n > 0,$$

and let the cluster set of the sequence on $|z| = 1$ on these arcs be S. Then if there is a point α omitted by the sequence, not belonging to S, and such that

$$\lim_{n\to\infty} f_n(0) = \alpha, \tag{4.11}$$

every point except α of the domain D containing α and bounded only by points of S is assumed by the sequence.

* The frontier points of a point set are the points every neighborhood of which contains a point both of the set and of its complement. If every frontier point of a domain belongs to a point set S, the domain will be said to be bounded only by points of S.

The fact that, in the case considered, a subset of S necessarily bounds a finite domain is not surprising in view of the information given by Theorem 3.3 about the oscillation of arc $[F_n(z)-\alpha]$ (where $F_n(z)$ is the Fatou boundary function of $f_n(z)$). The theorem will be proved first under the hypothesis that $f_n(z)$ is continuous on A_n.

Suppose that $\beta\neq\alpha$ were a point of D not assumed by the sequence. Then there would be a subsequence $\{f_{a_n}(z)\}$ omitting the value β. Let D' be a domain which, together with all its frontier points, is contained in D, and which contains the points α and β. Then $f_{a_n}(z)$ on A_{a_n} is outside D' for large values of n. Now if $\psi(w)$ is defined by

$$(4.12)\qquad \psi(w) = \Im \log\left(\frac{w-\alpha}{w-\beta}\right) = \operatorname{arc}(w-\alpha) - \operatorname{arc}(w-\beta),$$

it is seen that $\psi(w)$ is single-valued and continuous in the complementary set of D', choosing that branch for which $\psi(\infty)=0$. Then $\psi(w)$ must be bounded in the complement of D'. In particular, there is a number M such that

$$(4.13)\qquad |\psi[f_{a_n}(z)]| = |\operatorname{arc}[f_{a_n}(z)-\alpha] - \operatorname{arc}[f_{a_n}(z)-\beta]| \leqq M$$

for z on A_{a_n} and for values of n so large that $f_{a_n}(z)$ is outside D' on $A_{a_n}: n\geqq N$. Then if $E_{a_n}\subset A_{a_n}$ is a measurable point set on $|z|=1$,

$$(4.14)\qquad |O\{\operatorname{arc}[f_{a_n}(z)-\alpha], E_{a_n}\} - O\{\operatorname{arc}[f_{a_n}(z)-\beta], E_{a_n}\}| \leqq 2M,\ n\geqq N.$$

Now by Theorem 3.3, if $\liminf_{n\to\infty} mE_{a_n}>0$,

$$(4.15)\qquad \lim_{n\to\infty}\frac{\underline{B}\{|f_{a_n}(z)-\alpha|, E_{a_n}\}}{1+O\{\operatorname{arc}[f_{a_n}(z)-\alpha], E_{a_n}\}} = 0.$$

This implies that

$$(4.16)\qquad \lim_{n\to\infty} O\{\operatorname{arc}[f_{a_n}(z)-\alpha], E_{a_n}\} = +\infty,$$

since $\liminf_{n\to\infty}\underline{B}\{|f_{a_n}(z)-\alpha|, E_{a_n}\}$ is not less than the minimum of the distance from α to a point of S, which is positive. But (4.16) implies, by (4.14), that

$$(4.17)\qquad \lim_{n\to\infty}\frac{\underline{B}\{|f_{a_n}(z)-\beta|, E_{a_n}\}}{1+O\{\operatorname{arc}[f_{a_n}(z)-\beta], E_{a_n}\}} = 0,$$

since the denominator becomes infinite while the numerator is bounded uniformly for all values of n. By Theorem 3.3, (4.17) implies that $\lim_{n\to\infty} f_{a_n}(0)=\beta$, which is impossible since by hypothesis $\lim_{n\to\infty} f_{a_n}(0)=\alpha\neq\beta$. The hy-

pothesis that $\beta \neq \alpha$ was a point of D not assumed by the sequence $\{f_n(z)\}$ has thus led to a contradiction. The proof will now be given without the restriction that $f_n(z)$ be continuous on A_n. Let $\mathcal{J}_n(z)$ be the cluster boundary function of $f_n(z)$ on $|z|=1$. Let the arc A_n' have the same midpoint as A_n but be of half the length. Let $\mathcal{A}_n'$ be that arc on $|z|=r_n<1$ (where r_n will be determined below), which is cut off on $|z|=r_n$ by the sector of the unit circle intercepting A_n'. Then r_n can be so chosen that $1-r_n<1/n$ and that if $\mathfrak{z}$ is any point on $\mathcal{A}_n'$, there is a point z on A_n such that

$$|\mathcal{J}_n(z) - f_n(\mathfrak{z})| \leqq 1/n, \tag{4.18}$$

for some determination of $\mathcal{J}_n(z)$ at z. Now consider the sequence $\{f_n(r_nz)\}$. This sequence evidently has a subset S' of S as a cluster set on $|z|=1$ on the arcs $\{A_n'\}$. The function $f_n(r_nz)$ is continuous on A_n', the sequence omits the value α (not belonging to S') and $\lim_{n\to\infty} f_n(0)=\alpha$. Then by what has been proved already, the sequence assumes every value except α in the domain D' containing α and bounded only by points of S'. Since $S'\subset S$, $D'\supset D$, and the lemma is proved.

Lemma 4.2. *Let $\{f_n(z)\}$ be a sequence of functions meromorphic for $|z|<1$. Let $\{A_n\}$ be a sequence of arcs on $|z|=1$, and let the cluster set of the sequence in $|z|<1$ with respect to these arcs and on $|z|=1$ on these arcs be s and S, respectively. Then if there is a point α belonging to s but not to S and if α is omitted by the sequence $\{f_n(z)\}$, there is at most one other point in the domain D containing α and bounded only by points of S which is omitted by the sequence. If there are two points in D which are omitted by the sequence, no other point of the extended plane can be omitted by the sequence.*

(a) By hypothesis there is a subsequence $\{f_{a_n}(z)\}$ and a sequence $\{z_{a_n}\}$ of points in $|z|<1$, such that (2.05) is true and such that if A_{a_n} is transformed into A_{a_n}' by (2.09), (2.10) is true. We shall suppose that $f_n(z)\neq\alpha$, except for a finite number of values of n, that

$$\lim_{n\to\infty} f_n(0) = \alpha, \tag{4.21}$$

and that

$$\liminf_{n\to\infty} mA_n > 0. \tag{4.22}$$

If this is not true already, we can use the sequence $\{g_n(z)\}$, where

$$g_n(z) = f_{a_n}\left(\frac{z - z_{a_n}}{\bar{z}_{a_n}z-1}\right), \tag{4.23}$$

in conjunction with the arcs $\{B_n\}$ where $B_n = A_{a_n}'$. In this form the connection between this and the previous lemma is obvious.

(b) Suppose that besides α, a, b, c are also values omitted by the sequence $\{f_n(z)\}$, where α, a, b, c are supposed distinct and where a, b, c do not necessarily belong to D. We can suppose that a, b, c are 0, 1, ∞ respectively. For if this were not so we should prove the corresponding theorem for the sequence $\{h_n(z)\}$, which has 0, 1, ∞ as exceptional points:

$$h_n(z) = \frac{f_n(z) - a}{f_n(z) - b} \cdot \frac{c - b}{c - a} \tag{4.24}$$

if a, b, c are all finite. If one of the points is ∞, we can suppose that it is the point c, and we define $h_n(z)$ by

$$h_n(z) = \frac{f_n(z) - a}{f_n(z) - b}. \tag{4.25}$$

Let $\xi = \lambda_1(\xi')$ be a single-valued analytic function mapping $|\xi'| < 1$ on the extended ξ-plane less the points 0, 1, ∞ (the elliptic modular function defined in the circle instead of in the half-plane). The function $\lambda_1(\xi')$ maps $|\xi'| < 1$ in a one-to-one way on an infinitely many sheeted Riemann surface with branch points at 0, 1, ∞. Let $\lambda(\xi)$ be the inverse of $\lambda_1(\xi')$, and form the sequence $\{\phi_n(z)\}$ where

$$\phi_n(z) = \lambda[f_n(z)]. \tag{4.26}$$

Since $f_n(z)$ omits the values 0, 1, ∞, $\phi_n(z)$ can be taken as any one of an infinite set of single-valued analytic functions defined in $|z| < 1$, by the monodromic theorem, and $|\phi_n(z)| < 1$. Choose some determination of $\lambda(\alpha):\alpha'$. Then a branch of $\phi_n(z)$ can be chosen for each value of n so that

$$\lim_{n\to\infty} \phi_n(0) = \alpha'. \tag{4.27}$$

Since $f_n(z) \neq \alpha$, $\phi_n(z) \neq \alpha'$. Let the cluster set of the sequence $\{\phi_n(z)\}$ on $|z| = 1$ on the arcs $\{A_n\}$ be S'. If ξ' is a cluster value of the function $\phi_n(z)$ at a point of A_n, and if $|\xi'| < 1$, $\lambda_1(\xi')$ is a cluster value of $f_n(z)$ at that point of A_n. Then if ξ' is a point of S' and if $|\xi'| < 1$, $\lambda_1(\xi')$ is a point of S. Then α' cannot belong to S' or α would belong to S, since $|\alpha'| < 1$. Let D' be the domain containing α' and bounded only by points of S'. Then by Lemma 4.1 every point in D' except α' is assumed by the sequence $\{\phi_n(z)\}$.

Now suppose that β is a point of D, $\beta \neq \alpha$, 0, 1, ∞. Let J be a Jordan arc joining α to β and lying wholly in D. It can be so chosen that it does not pass through 0, 1 or ∞. Choose the branch of $\lambda(\xi)$ for which $\lambda(\alpha) = \alpha'$ and using

that branch determine J', the image of J in the ξ'-plane. Then we shall prove that J' lies wholly in D'. One end point, α', belongs to D'. If J' is not wholly in D', there is a point of J' on the boundary S' of D'. We can suppose that ξ' is the first such point on J', tracing J' from α'. If $|\xi'|<1$, $\xi=\lambda_1(\xi')$ belongs to S, as was noted above. But ξ is on J, belongs to D, and so cannot belong to S. If $|\xi'|=1$, the arc J must spiral infinitely often about 0, 1, or ∞, since it remains at positive distance from the first two and remains in some circle about the origin. But this is impossible, since J is a Jordan arc. Then no point on J' can be on the boundary of D', so J' lies entirely in D'. This means that $\beta'=\lambda(\beta)$ belongs to D' and is therefore assumed by the sequence $\{\phi_n(z)\}$. Then β is assumed by the sequence $\{f_n(z)\}$.

It has thus been proved that if three points a, b, c are exceptional, besides α, every point of D save α and the points of a, b, c belonging to D, is assumed by the sequence (a subsequence of the original sequence).

(c) The result of (b) will now be sharpened. Suppose that $\beta\neq\alpha$ belongs to D and is an exceptional value of the sequence $\{f_n(z)\}$ for which (4.21) and (4.22) are true. Suppose that γ is a third exceptional value of the sequence, not necessarily in D. We can suppose that $\alpha=0$, $\beta=\infty$, $\gamma=1$. Consider the sequence $\{\psi_n(z)\}$ where

$$\psi_n(z) = (f_n(z))^{1/3}. \tag{4.28}$$

Choosing any one of three branches, $\psi_n(z)$ is single-valued and analytic in $|z|<1$, by the monodromic theorem, since $f_n(z)\neq 0, \infty$. Moreover the sequence $\{\psi_n(z)\}$ has the exceptional values

$$\alpha_1 = 0,\ \beta_1 = \infty,\ a = e^{2\pi i/3},\ b = a^2,\ c = 1,$$

and

$$\lim_{n\to\infty} \psi_n(0) = 0. \tag{4.29}$$

Let S_1 be the cluster set of the sequence $\{\psi_n(z)\}$ on $|z|=1$ on the arcs $\{A_n\}$. If ξ_1 is a point of S_1, $\xi=\xi_1^3$ is a point of S. The point $\alpha_1=0$ therefore does not belong to S_1 or $\alpha=0$ would belong to S. Let D_1 be the domain containing α_1 and bounded only by points of S_1. It will be shown that D_1 contains β_1. Let J be a Jordan arc in D with end points $\alpha=0$, $\beta=\infty$.* A point ξ_0 can be chosen on J, so near $\alpha=0$ that each determination of $\xi^{1/3}$ lies in D_1—since D_1 includes some neighborhood of the origin. Then continue one determination of $\xi^{1/3}$ from ξ_0 to β along J, thus determining a Jordan arc J_1, which we shall prove lies entirely in D_1. For if it did not, a point of J_1 would also be a point of S_1,

* This means that to J corresponds a Jordan arc on the sphere.

the boundary of D_1, which would imply that a point of J was a point of S. Since this is not true, $\beta_1=\infty$ must be a point of D_1. Now the sequence $\{\psi_n(z)\}$ omits three values a, b, c besides α_1, so that every point in D_1 (not one of these values) must be assumed by the sequence, as was proved in (b). Therefore $\beta_1=\infty$ must be assumed. The sequence $\{f_n(z)\}$ must therefore assume the value $\beta=\infty$, contrary to hypothesis.

It has thus been shown that if there is an exceptional value in D besides α, no other point in the extended plane is omitted by the sequence. This proves the lemma.

LEMMA 4.3. *Let A be an arc of $|z|=1$ and let $f(z)$ be a bounded function, analytic for z in the segment $\mathfrak{S}(A)$ of the unit circle, with Fatou boundary function $F(z)$ on A. Suppose that*

$$|f(z)| \leqq K \text{ in } \mathfrak{S}(A) \tag{4.31}$$

and that

$$|F(z)| \leqq k < K \text{ on } A. \tag{4.32}$$

Then if $k'>k$,

$$|f(z)| \leqq k' \text{ in } \mathfrak{S}(A') \tag{4.33}$$

where $A' \subset A$ and where mA' is a function of mA, k'/k, K/k only,

$$mA' = \tau(mA, k'/k, K/k), \tag{4.34}$$

$\tau>0$ increasing with mA, k'/k, k/K.

This lemma is easily proved using the maximum principle.*

LEMMA 4.4. *Let $\{f_n(z)\}$ be a sequence of functions meromorphic for $|z|<1$. Let the cluster set and the strong cluster set of the sequence in $|z|<1$ with respect to a set of arcs $\{A_n\}$ on $|z|=1$ be s, s_1 respectively, and let the cluster set of the sequence on $|z|=1$ on these arcs be S.*

(a) *$S \subset s_1 \subset s$.*

(b) *The points of s not belonging to S form an open set† consisting of non-overlapping domains every one of which has at least one frontier point belonging to S.*

(c) *If one point α of one of these domains belongs to s_1, every point of the domain containing α and bounded only by points of S belongs to s_1.*

---

* Cf. for example the proof of a related fact obtained by W. Seidel, these Transactions, vol. 34 (1932), pp. 3–4.

† This set may be empty.

(a) Let P belong to S. By hypothesis there is a sequence of points $\{P_{a_n}\}$, where P_{a_n} is a point of A_{a_n}, such that $\lim_{n\to\infty} \mathcal{J}_{a_n}(P_{a_n})=P$, $\mathcal{J}_{a_n}(P_{a_n})$ representing one of the values of $\mathcal{J}_{a_n}(z)$, the cluster boundary function of $f_{a_n}(z)$, at P_{a_n}. We can suppose that $\liminf_{n\to\infty} mA_{a_n}>0$. There is a sequence of points in $|z|<1$ approaching P_{a_n}, on which $f_{a_n}(z)$ approaches $\mathcal{J}_{a_n}(P_{a_n})$. Then let z_{a_n} be one of these points so close to P_{a_n} that

$$|\mathcal{J}_{a_n}(P_{a_n}) - f_{a_n}(z_{a_n})| < 1/n$$

and that

$$z' = \frac{z - z_{a_n}}{\bar{z}_{a_n}z - 1}$$

transforms A_{a_n} into an arc of length not less than $2\pi-1/n$. The existence of the sequence $\{z_{a_n}\}$ is the condition that P belong to s_1. Then $S\subset s_1$, and, by definition, $s_1\subset s$.

(b) The first part of (b) is equivalent to the statement that the frontier points of s which belong to s also belong to S. Suppose the contrary, that there is a point α, a frontier point of s belonging to s but not to S. We can suppose that α is finite, substituting the sequence $\{1/f_n(z)\}$ for $\{f_n(z)\}$ if $\alpha=\infty$. Making, if necessary, linear transformations taking $|z|<1$ into itself for all values of n, we can suppose that

$$\lim_{n\to\infty} f_{a_n}(0) = \alpha \tag{4.41}$$

and that

$$\liminf_{n\to\infty} mA_{a_n} > 0. \tag{4.42}$$

The sequence $\{f_{a_n}(z)\}$ is normal. For otherwise there would be a point z_0 with the property that in any neighborhood of z_0 at most two values are omitted by the sequence.* This would mean that every value of the plane belonged to s, contradicting the fact that α is a frontier point of s. If there were a limit function $f(z)\not\equiv\alpha$, $f(0)=\alpha$ necessarily and $f(z)$ would assume every value in some neighborhood of α, for $|z|<\frac{1}{2}$. Then the cluster set of $\{f_{a_n}(z)\}$ in $|z|<\frac{1}{2}$, which is a subset of s, would include this neighborhood of α contrary to the hypothesis that α was a frontier point of s. The sequence $\{f_{a_n}(z)\}$ is thus a normal family with the single limit function α, which implies, by an argument similar to that used in the proof of Lemma 3.1, that

$$\lim_{n\to\infty} f_{a_n}(z) = \alpha \tag{4.43}$$

uniformly in every closed subregion of $|z|<1$.

* P. Montel, *Leçons sur les Familles Normales*, Paris, 1927, p. 126.

Let every point of S be at distance greater than $3d>0$ from α. Then

$$(4.44)\qquad |\mathcal{J}_{a_n}(z)-\alpha|>2d \text{ on } A_{a_n}$$

for large values of $n:n\geqq N$, and every determination of $\mathcal{J}_{a_n}(z)$. Since α is a frontier point of s there is a point β not belonging to s such that

$$(4.45)\qquad |\alpha-\beta|<d.$$

Then

$$(4.46)\qquad |F_{a_n}(z)-\beta|>d \text{ on } A_{a_n},\ n\geqq N.$$

There is no sequence of points $\{\xi_{b_n}\}$ such that ξ_{b_n} belongs to $\mathfrak{S}(A_{b_n})$ and such that $\lim_{n\to\infty}f_{b_n}(\xi_{b_n})-\beta=0$, where $\{f_{b_n}(z)\}$ is a subsequence of $\{f_{a_n}(z)\}$. For then β would have to be a point of s, contrary to hypothesis. Then there is a number K with the property that

$$(4.47)\qquad 1/|f_{a_n}(z)-\beta|\leqq K \text{ in } \mathfrak{S}(A_{a_n})$$

for large values of n: $n\geqq N_1$. But then if $d'<d$, by the previous lemma there is a sequence of arcs $\{A_{a_n}'\}$, $A_{a_n}'\subset A_{a_n}$, such that $\liminf_{n\to\infty} mA_{a_n}'>0$ and such that

$$(4.48)\qquad 1/|f_{a_n}(z)-\beta|\leqq 1/d' \text{ in } \mathfrak{S}(A_{a_n}),\ n\geqq N, N_1.$$

Then by (4.3)

$$(4.49)\qquad 1/|\alpha-\beta|\leqq 1/d'.$$

Since d' was arbitrary, $d'<d$, (4.49) implies that $|\alpha-\beta|\geqq d$ which contradicts (4.45). The hypothesis that there was a frontier point of s belonging to s but not to S has thus led to a contradiction.

The points common to s and the complement of S thus form an open set. This set is the sum of non-overlapping domains. At least one frontier point of each domain belongs to S. For if D is one of these domains and if α is a point of D, we can suppose that (4.41) and (4.42) are true. Consider the set E of all limit values of sequences of the form $\{f_{a_n}(\xi_{a_n})\}$ where ξ_{a_n} is a point of $|z|<1$ on the radius from $z=0$ to Q_{a_n}, the midpoint of A_{a_n}. This set E, a subset of s, is readily seen to be closed and connected and to contain α and also at least one point of S (in fact one of the limit values of the sequence $\{\mathcal{J}_{a_n}(Q_{a_n})\}$). By a well known theorem, since both a point in D and a point not in D belong to E, a frontier point of D must belong to E. We shall prove that this point P belongs to S, thus completing the proof of (b). If P did not belong to S, it would be a frontier point of s which belonged to s, since $E\subset s$. This is impossible by what has been proved already. Then P belongs to S.

(c) Statement (c) is equivalent to the statement that the frontier points

of s_1 (which belong to s_1 since s_1 is closed) are points of S. The proof is similar to that of (b).

We now combine Lemma 4.4 with Lemma 4.2 to get the final result of this section.

THEOREM 4.2. *Let $\{f_n(z)\}$ be a sequence of functions meromorphic for $|z|<1$. Let the cluster set and the strong cluster set of the sequence in $|z|<1$ with respect to a set of arcs $\{A_n\}$ on $|z|=1$ be s and s_1 respectively and let S be the cluster set of the sequence on $|z|=1$ on these arcs. Let there be a point α belonging to s but not to S.*

(a) *Suppose that no point of the domain D containing α and bounded only by points of S belongs to s_1. Then if one point of the set $s\cdot D$, consisting of non-overlapping domains each with at least one frontier point belonging to S, is omitted by the sequence $\{f_n(z)\}$, only one other point of the extended plane can be omitted and every point of the extended plane belongs to s.*

(b) *If a point of s_1 is in D, $D\subset s_1$, and at most two points of D are omitted by the sequence $\{f_n(z)\}$. If two points of D are omitted no other point of the extended plane is omitted and every point of the extended plane belongs to s.*

(a) In (a) if a point α of $s\cdot D$, which was described in Lemma 4.4 (b), is exceptional to the sequence $\{f_n(z)\}$, α cannot be a convergence value of the sequence with respect to the arcs $\{A_n\}$, or α would belong to s_1 (cf. §2). Then if we suppose, as we can, that a subsequence $\{f_{a_n}(z)\}$ exists for which

$$\lim_{n\to\infty} f_{a_n}(0)=\alpha,\ \liminf_{n\to\infty} mA_{a_n}>0,$$

the sequence $\{f_{a_n}(z)\}$ cannot be normal or α would be a convergence value by an argument used in the proof of Theorem 3.2 (a). Then by a theorem used above there must be a point in $|z|<1$ in every neighborhood of which the sequence $\{f_{a_n}(z)\}$ can have at most two exceptional values, i.e. one besides α. This proves (a) completely.

(b) If a point α of s_1 is in D, $D\subset s_1\subset s$ by Lemma 4.4 (c). Then any exceptional value of the sequence in D is an exceptional cluster value with respect to the arcs $\{A_n\}$, and Lemma 4.2 can be applied. There only remains the proof that if two points of D are exceptional, s is the entire extended plane. We can suppose that there is a subsequence $\{f_{a_n}(z)\}$ such that

$$\lim_{n\to\infty} f_{a_n}(0)=\alpha, \text{ and } \lim_{n\to\infty} mA_{a_n}=2\pi.$$

We consider $f_{a_n}(z)$ for n so large that $mA_{a_n}\geqq 3\pi/2$ considering the values of $f_{a_n}(z)$ for z in the interior of a segment $\mathfrak{S}(A_{a_n}')$ determined by a subarc A_{a_n}' of A_{a_n} of length $3\pi/2$. Then it is immediate that if β is arbitrary except that β is not one of the two values in D exceptional to the sequence $\{f_n(z)\}$ by hypothesis, $f_{a_n}(z)-\beta=0$ has a root in $\mathfrak{S}(A_{a_n}')$ for an infinite set of

values of n (proof by mapping $\mathfrak{S}(A_{a_n}')$ on a new unit circle and applying Lemma 4.2). Then β is a point of s, as was to be proved.

5. The neighborhood properties of the boundary function of a bounded analytic function

Let $f(z)$ be a bounded analytic function, defined for $|z|<1$, with Fatou boundary function $F(z)$. We shall discuss the following two questions. Let $P:e^{it}$ be a point on $|z|=1$. What are necessary and sufficient conditions on $F(z)$ in a neighborhood of P that $F(z)$ be defined at P: $\lim_{r\to 1} f(re^{it})=F(P)$? What are necessary and sufficient conditions on $F(z)$ in a neighborhood of P that $f(z)$ have the cluster value α at P? The latter case can be divided into two parts, according as α is or is not a non-tangential cluster value. The most stress in this section will be laid on conditions which are both necessary and sufficient and for this reason and for reasons of simplicity the sufficient conditions will not be stated with the full generality possible.

Theorem 5.1. *Let $f(z)$ be a bounded function analytic for $|z|<1$ with Fatou boundary function $F(z)$, $|F(z)|\leqq 1$. Let P be a point on $|z|=1$.*

(a) *If $\lim_{z\to P}|f(z)|=1$ when z approaches P on a continuous curve C, lying on one side of some chord through P, $|F(z)|$ is approximately continuous at P on that side, if $|F(P)|$ is defined as 1. In particular, if C is a non-tangential path, $\lim_{z\to P}|f(z)|=1$ when z approaches P on every non-tangential path and $|F(z)|$ is approximately continuous at P if $|F(P)|$ is defined as 1.*

(b) *If $|f(z)|$ has the cluster value 1 at P, $E(|F(z)|\geqq 1-\epsilon)$* is metrically dense at P for all positive values of ϵ. If 1 is a cluster value on some non-tangential path it is a cluster value on every continuous non-tangential or tangential curve to P and $|F(z)|$ is quasi-approximately continuous at P with limit value 1 there.*

(a) It is convenient to prove the second part of (a) first. Define $f_1(\xi)$, analytic in the upper half-plane with boundary function $F_1(\xi)$ by

$$f_1(\xi)=f\left(\frac{1+i\xi}{1-i\xi}\right),\quad F_1(\xi)=F\left(\frac{1+i\xi}{1-i\xi}\right). \tag{5.101}$$

Using Lemma 1.1, we see that it is sufficient to prove the result corresponding to the second part of (a) for $f_1(\xi)$ and its boundary function $F_1(\xi)$. We can suppose that P is the point $|z|=1$. Non-tangential paths to a point of the

* If a function $F(z)$ is defined almost everywhere on $|z|=1$, it will be convenient to denote the set of points on $|z|=1$ at which $F(z)$ satisfies a given inequality by $E(\quad)$, where the inequality is enclosed by the parentheses.

real axis are defined as paths which remain within some angle with vertex at the point whose sides are rays in the half-plane under consideration.

By hypothesis, then, $\lim_{\xi\to 0} |f_1(\xi)| = 1$, when ξ approaches $\xi = 0$ on C_1, a non-tangential continuous curve. C_1 is included in the angle determined by two rays, L', L'', meeting at $O: \xi = 0$. We can suppose that L', L'' are symmetric in the imaginary axis. Let R_λ be the interior of the rectangle having one side, of length λ, on the real axis and opposite side with end points on L' and L''. The rectangle is symmetric in the imaginary axis. Let P_λ be the intersection of the diagonals of R_λ and let $\phi(w)$ be the function mapping $|w| < 1$ in a one-to-one and conformal way on R_1, so that $\phi(0) = P_1$ and so that $\phi'(0)$ is real and positive.* We consider the family $\{g_\lambda(w)\}$ where $g_\lambda(w) = f_1[\lambda\phi(w)]$, $0 < \lambda \leqq 1$. The function $g_\lambda(w)$ takes on those values in $|w| < 1$ which $f_1(\xi)$ takes on in R_λ. Let $G_\lambda(w)$ be the Fatou boundary function of $g_\lambda(w)$.

(i) $\lim_{\lambda\to 0} |g_\lambda(w)| = 1$ uniformly in every closed subregion of $|w| < 1$. For there is a value of $\rho < 1$ such that $|w| = \rho$ corresponds to a simple closed analytic curve J in R_1 (by means of the transformation $\xi = \phi(w)$) which intersects both L' and L'' and therefore C_1.† Then for each positive value of λ, there is a point w_λ, $|w_\lambda| = \rho$, such that $\lambda\phi(w_\lambda)$ is a point of C_1. Then

$$\lim_{\lambda\to 0} |g_\lambda(w_\lambda)| = 1, \quad |w_\lambda| = \rho < 1, \tag{5.102}$$

and Lemma 3.1 proves the desired statement.‡

(ii) It follows from (i) that $\lim_{\xi\to 0} |f_1(\xi)| = 1$ uniformly in the angle considered.

(iii) It follows from (5.102), by Lemma 3.1, that the family $\{|G_\lambda(w)|\}$ converges in measure to 1 ($\lambda\to 0$). If $E_\lambda(\epsilon)$ is the set of those points on $|w| = 1$ for which $|G_\lambda(w)| \leqq 1-\epsilon$, $E_\lambda(\epsilon)$ corresponds to a set $E_\lambda'(\epsilon)$ on R_λ by the transformation $\xi = \lambda\phi(w)$ which is continuous on $|w| = 1$. The set $E_\lambda'(\epsilon)$ consists of those points on the perimeter of R_λ for which $|F_1(\xi)| \leqq 1-\epsilon$. The set $E_\lambda(\epsilon)$ corresponds to a set of measure $mE_\lambda'(\epsilon)/\lambda$ on R_1 by the transformation $\xi = \phi(w)$. We have

* The function $\phi(w)$ can be given by means of elliptic integrals: see for example W. F. Osgood, *Lehrbuch der Funktionentheorie*, 5th edition, 1928, p. 437.

† This follows from the fact that if $\{w_n\}$ is a sequence of points in $|w| < 1$ and if

$$\lim_{n\to\infty} |w_n| = 1,$$

the sequence $\{\phi(w_n)\}$ has no limit point in R_1, for which see, for example, L. Bieberbach, *Lehrbuch der Funktionentheorie*, 2d edition, vol. 2, 1931, pp. 24–25.

‡ The theorems of the preceding section can be stated for a family of functions depending on a parameter running through a continuous set of values to a limiting value just as well as for a family depending on the parameter n taking on only integral values.

(5.103) $$\lim_{\lambda\to 0} mE_\lambda(\epsilon) = 0,$$

which implies that

(5.104) $$\lim_{\lambda\to 0} mE'_\lambda(\epsilon)/\lambda = 0^*.$$

But the measure of the set $\mathcal{E}_\lambda(\epsilon)$ of those points on the real axis on the interval $|\xi| \leqq \lambda/2$ for which $|F_1(\xi)| \leqq 1-\epsilon$ being less than $mE'_\lambda(\epsilon)$, it follows that

(5.105) $$\lim_{\lambda\to 0} m\mathcal{E}_\lambda(\epsilon)/\lambda = 0,$$

i.e. that $\mathcal{E}_\lambda(\epsilon)$ has density 0 at $\xi=0$. Since $\epsilon>0$ was arbitrary, it has been proved that $|F_1(\xi)|$ is approximately continuous at $\xi=0$, if $|F_1(0)|$ is defined as 1.

To prove the first part of the statement (a) of the theorem we change the domain of definition of the function again. Define $f_2(\xi)$ analytic in the first quadrant, with boundary function $F_2(\xi)$, by

(5.106) $$f_2(\xi) = f\left(\frac{1+i\xi^2}{1-i\xi^2}\right), \quad F_2(\xi) = F_2\left(\frac{1+i\xi^2}{1-i\xi^2}\right).$$

By the corollary to Lemma 1.1 it is sufficient to prove the results desired for the function $f_2(\xi)$ and its boundary function $F_2(\xi)$ (we suppose that P is the point $z=1$, and the point in question in the ξ-plane is then the origin). Suppose then that $\lim_{\xi\to 0} |f_2(\xi)| = 1$ when ξ approaches $O: \xi=0$ on a continuous curve C_2 which lies in the angle formed by a ray L in the first quadrant with one of the rays which bound the first quadrant. We can suppose that C_2 lies in the angle between L and the (positive) real axis. Let the angle between L and the positive real axis be θ and let L' be a ray through O into the first quadrant making an angle $\theta'>\theta$, $\theta'<\pi/2$, with the positive real axis. Let P_1 be a point of C_2 and let R_1 be the rectangle whose diagonals L_1', L_1'' intersect at P_1, and which has two vertices Q_1, Q_2 on the real axis. We suppose R_1 so chosen that $L_1' \| L'$. Then Q_1, Q_2 are both on the positive real axis; take $OQ_1>OQ_2$. The ray through Q_2 parallel to L_1'' must meet C_2 in at least one point. Let P_2 be the intersection of the ray and C_2 which is nearest Q_2, and let R_2 be the rectangle whose diagonals L_2', L_2'' intersect at P_2, $L_2' \| L_1'$, $L_2'' \| L_1''$, and with vertices Q_2, Q_3 on the real axis. The point Q_3 must be on the positive real axis since $L_2' \| L'$. In this way we get a sequence of rectangles $R_1, R_2, \cdots$ whose diagonals intersect at $P_1, P_2, \cdots$ and a sequence

* This can be proved most readily using the fact that $\phi(w)$ on $|w|=1$ is continuous and has a continuous derivative except at four points in the neighborhood of which $\phi(w)$ has the same character as $w^{1/2}$ at $w=0$.

of vertices $Q_1, Q_2, \cdots$ on the positive real axis, $OQ_1 > OQ_2 \cdots$. The monotone sequence $\{Q_n\}$ has a unique limiting point which must be O, for if it were not O, the sequence $\{P_n\}$ would have as unique limit that same point on the positive real axis, which is impossible since P_n is on C_2 for all values of n. Let I_n be the closed interval with end points Q_n, Q_{n+1}. Then it is easily seen that

$$(5.107) \qquad \frac{mI_n}{OQ_{n+1}} = \frac{mI_n}{\sum_{j+1}^{\infty} mI_j} \leqq \frac{2 \tan \theta}{\tan \theta' - \tan \theta} = M.$$

Let $\phi_n(w)$ map $|w| < 1$ in a one-to-one and conformal way on the interior of R_n so that $\phi_n(0) = P_n$, $\phi_n'(0) > 0$. Then

$$(5.108) \qquad \phi_n'(w) = \frac{mI_n}{mI_j} \phi_j'(w).$$

Consider the sequence $\{f_2[\phi_n(w)]\}$. Then

$$(5.109) \qquad |f_2[\phi_n(w)]| \leqq 1, \quad \lim_{n\to\infty} |f_2[\phi_n(0)]| = 1.$$

Therefore, by Lemma 3.1, the sequence of the absolute values of the boundary functions is convergent in measure to 1 on $|z| = 1$. Repeating an argument used above, if $E(\epsilon)$ is the set of those points on the positive real axis for which $|F_2(\xi)| \leqq 1 - \epsilon$, $\epsilon > 0$,

$$(5.110) \qquad \lim_{n\to\infty} \frac{m[I_n \cdot E(\epsilon)]}{mI_n} = 0.$$

Now let I be a variable closed interval with one end point at O and the other on the positive real axis. Let $\lambda = \lambda(I)$ be the smallest value of j for which Q_j belongs to I. Then

$$(5.111) \qquad \sum_{j=\lambda}^{\infty} mI_j \leqq mI \leqq \sum_{j=\lambda-1}^{\infty} mI_j, \quad mI_n \leqq M \sum_{j=n+1}^{\infty} mI_j,$$

and

$$(5.112) \qquad m[I \cdot E(\epsilon)] \leqq \sum_{j=\lambda-1}^{\infty} m[I_j \cdot E(\epsilon)].$$

Choose $\delta > 0$ so that

$$(5.113) \qquad \frac{m[I_j \cdot E(\epsilon)]}{mI_j} \leqq \frac{\eta}{1+M} \text{ for } j \geqq \lambda - 1, \text{ if } mI \leqq \delta,$$

for some fixed positive number η. Then

$$(5.114)\qquad \frac{m[I\cdot E(\epsilon)]}{mI} \leqq \frac{\eta}{(1+M)mI}\sum_{j=\lambda}^{\infty} mI_j + \frac{\eta}{1+M}\,\frac{mI_{\lambda-1}}{mI} \leqq \eta.$$

Since η was an arbitrary positive number, $|F_2(\xi)|$ must be approximately continuous at O on the positive real axis, if $|F_2(0)|$ is defined as 1, as was to be proved.

(b) If $|f(z)|$ has the cluster value 1 at P (which we suppose to be the point $z=1$) we go again to the function $f_1(\xi)$ defined by (5.101). We have a sequence of points $\{P_n\}$, $P_n\to O: \xi=0$, such that

$$(5.115)\qquad \lim_{n\to\infty} |f_1(P_n)| = 1.$$

Let R_1 be a rectangle with one side on the positive real axis and whose diagonals intersect at P_1. Let R_n, $n>1$, be the rectangle with one side on the real axis, whose diagonals are parallel to those of R_1 and intersect at P_n. The rectangles are all similar and by considering $f_1(\xi)$ defined in R_n, we find by reasoning similar to that used above that there is a sequence of intervals on the real axis (the bases of the rectangles) such that, with self-explanatory notation,

$$(5.116)\qquad \lim_{n\to\infty} \frac{mI_n\cdot E(|F_1(\xi)| \leqq 1-\epsilon)}{mI_n} = 0,$$

which shows that $E(|F_1(\xi)| \geqq 1-\epsilon)$ is metrically dense at $\xi=0$ for all values of $\epsilon>0$, implying the same for $E(|F(z)| \geqq 1-\epsilon)$ at P on $|z|=1$.

If the sequence $\{\xi_n\}$ is non-tangential, the first proof given in (a) can be used, choosing a suitable sequence from the family $\{g_\lambda(w)\}$, to show that 1 is a cluster value of $|f(z)|$ on every continuous non-tangential path to P and that $|F(z)|$ is quasi-approximately continuous at P with limit value 1 there. There remains the proof that 1 is a cluster value on a continuous curve C which is tangent to $|z|=1$ at P. It is not difficult to reduce this to the results already proved, by the use of conformal mapping, and the details will not be given. If C is an arc of a circle, the preceding results show that $|f(z)|$ will be even quasi-approximately continuous on C at P, with limit value 1 there.

THEOREM 5.2. *Let $f(z)$ be a bounded function analytic for $|z|<1$ with Fatou boundary function $F(z)$, $|F(z)|\leqq 1$. Let P be a point on $|z|=1$.*

(a) *If $F(z)$ is defined at P and if $|F(P)|=1$, $F(z)$ is approximately continuous at P.**

* The converse is empty, strictly speaking, since in the definition of approximate continuity at P, $F(z)$ is supposed defined at P; cf. however Theorem 3 of the previous paper referred to above, and a note below.

(b) *A necessary and sufficient condition that $f(z)$ have the cluster value α at P, if $|\alpha| = 1$, is that $E(|F(z)-\alpha| \leqq \epsilon)$ be metrically dense at P for all positive values of ϵ. A necessary and sufficient condition that $f(z)$ have α as a non-tangential cluster value at P if $|\alpha| = 1$ is that $F(z)$ be quasi-approximately continuous at P with limit value α there. If α is a non-tangential cluster value at P, $|\alpha| = 1$, it is a cluster value on every continuous non-tangential or tangential path to P.*

(c) *In* (a) *if $w=f(z)$, for z on some continuous non-tangential curve to P, defines a curve in the w-plane more closely tangent to $|w| = 1$ at $w=F(P)$ than any circle C_ρ of radius $\rho<1$, in $|w| \leqq 1$ and tangent to $|w| = 1$ at $w=F(P)$, the same is true for the curve defined by $w=f(z)$ when z is on any non-tangential curve to P and the metric density of the set E_ρ of those points at which $F(z)$ is outside C_ρ is 1 for all values of $\rho<1$. In* (b) *if α is a cluster value at P:*

$$\lim_{n\to\infty} f(z_n) = \alpha, \quad \lim_{n\to\infty} z_n = P, \quad |\alpha| = 1, \tag{5.21}$$

and if the sequence $\{w_n\}$, $w_n = f(z_n)$, is more closely tangent to $|w| = 1$ at $w=\alpha$ than any circle C_ρ, $\rho<1$, the set E_ρ is metrically dense at P for all values of $\rho<1$. If the sequence $\{z_n\}$ in (5.21) *is non-tangential and if the sequence $\{w_n\}$ has the same property as above, E_ρ has upper mean metric density* 1 *at P for all values of $\rho<1$.*

The statements (a), (b) can be proved by an argument similar to that in the previous theorem, referring the result back to Theorem 3.1. The necessary conditions are simply Theorem 5.1 applied to

$$e^{[f(z)/\alpha]-1}.$$

We note that there are not two cases in (a) as there were in Theorem 5.1 because by a theorem of Lindelöf,* if $f(z)$ has a unique limit on a continuous curve to P, $f(z)$ will have that same limit on every non-tangential path.

The sufficient conditions are independent of the fact that $|\alpha| = 1$ (cf. Theorem 3.1).†

The statement (c) can be deduced from the corollary to Theorem 3.1 or (a), (b) of Theorem 5.1 can be applied to the function

$$e^{(f(z)+1)/(f(z)-1)}.$$

This result is a sort of complement to a well known theorem of Julia, Wolff and Carathéodory.‡

* E. Lindelöf, Acta Societatis Scientiarum Fennicae, vol. 46 (1915), No. 4, p. 10.

† By using the criterion for convergence of a sequence given in the generalization of Theorem 3.1 (b), we can deduce again Theorem 3 of the previous paper, referred to above.

‡ See for example L. Bieberbach, *Lehrbuch der Funktionentheorie*, 2d edition, vol. 2, 1931, pp. 112–121.

THEOREM 5.3. *Let $f(z)$ be a bounded function, analytic for $|z|<1$ with Fatou boundary function $F(z)$. Suppose that $f(z)\neq\alpha$ in some neighborhood of a point P on $|z|=1$.*

(a) *A necessary and sufficient condition that*

$$\lim 1/\log\,[f(z)-\alpha] = 0^*$$

when z approaches P on continuous non-tangential paths is that $1/\log\,[F(z)-\alpha]$ be approximately continuous at P if $1/\log\,[F(P)-\alpha]$ is defined as 0.

(b) *A necessary and sufficient condition that $1/\log\,[f(z)-\alpha]$ have the cluster value 0 at P is that $E\{|\log\,[F(z)-\alpha]|\geqq M\}$ be metrically dense at P for all values of M. A necessary and sufficient condition that $1/\log\,[f(z)-\alpha]$ have the non-tangential cluster value 0 at P is that $1/\log\,[F(z)-\alpha]$ be quasi-approximately continuous at P with limit value 0 there. If 0 is a non-tangential cluster value it is a cluster value on every continuous non-tangential or tangential path to P.*

This theorem can be deduced by means of Theorem 3.2 or by applying Theorem 5.2 to the function

$$\frac{\log f(z)+1}{\log f(z)-1}$$

where we suppose that $\alpha=0$ and that $|f(z)|<1$ (cf. the proof of Theorem 3.2).

THEOREM 5.4. *Let $f(z)$ be a bounded function analytic for $|z|<1$, with Fatou boundary function $F(z)$. Suppose that $f(z)\neq\alpha$ in some neighborhood of a point P on $|z|=1$.*

(a) *A necessary and sufficient condition that $F(z)$ be defined at P and that $F(P)=\alpha$ is that*

$$\lim_{n\to\infty}\frac{\underline{B}\{|F(z)-\alpha|,E_n\}}{1+O\{\operatorname{arc}\,[F(z)-\alpha],E_n\}}=0 \tag{5.41}$$

for every sequence $\{E_n\}$ of measurable point sets on $|z|=1$ with the property that there exists a sequence of arcs $\{A_n\}$ on $|z|=1$ with midpoint P such that

$$E_n\subset A_n,\quad \lim_{n\to\infty} mA_n=0,\quad \liminf_{n\to\infty}\frac{mE_n}{mA_n}>0. \tag{5.42}$$

(b) *A necessary and sufficient condition that $f(z)$ have α as a cluster value at P is that there be a sequence of arcs $\{A_n\}$ on $|z|=1$ whose end points approach P with the property that*

---

* Cf. the note on p. 427.

$$\lim_{n\to\infty} \frac{\underline{B}\{\,|F(z)-\alpha|,\, E\cdot A_n\}}{1+O\{\text{arc}\,[F(z)-\alpha],\, E\cdot A_n\}} = 0 \tag{5.43}$$

for every measurable set E such that

$$\liminf_{n\to\infty} \frac{mE\cdot A_n}{mA_n} > 0. \tag{5.44}$$

(c) *In* (b) *the conditions are necessary and sufficient that* $f(z)$ *have* α *as a non-tangential cluster value if the arcs* $\{A_n\}$ *are all taken to have the midpoint* P.

The corresponding statement for $f(z)$ defined in a half-plane is obvious, and it is convenient to prove it in this case. This is equivalent to proving the theorem as stated, as is shown by a slight extension of Lemma 1.1. The theorem is then easily deduced from Theorem 3.3 by considering $f(z)$ (defined in the half-plane $\mathfrak{J}(z)>0$) in suitable rectangles with bases on the real axis. The discussion is analogous to that in the proof of Theorem 5.1.

Corollary. *Theorem* 5.3 *gives necessary and sufficient conditions* (a) *that* $F(z)$ *be defined at* P, (b) *that* $f(z)$ *have a given cluster value at* P, (c) *that* $f(z)$ *have a given non-tangential cluster value at* P, *if* $f(z)$ *is supposed univalent.**

For a univalent function $f(z)$ defined for $|z|<1$ does not assume a cluster value at a point P on $|z|=1$ in some neighborhood of P.

The problem set at the beginning of the section has thus been solved in a special case. If $f(z)$ is bounded and analytic in the interior of the unit circle, with Fatou boundary function $F(z)$, necessary and sufficient conditions have been found on $F(z)$ in a neighborhood of a point P on $|z|=1$ that $F(z)$ be defined at P, and that $f(z)$ have the cluster value α at P, if $f(z)\neq F(P)$, $f(z)\neq\alpha$, respectively, in some neighborhood of P. In the first case the conditions need be modified only slightly if $f(z)=F(P)$ only at points of $|z|<1$ on one side of some chord through P; we need only consider $F(z)$ on one side of P on $|z|=1$. In both cases, by using Theorem 3.4, the general case can be solved, but the statement becomes so complicated that it is of no interest.

6. The neighborhood properties of the cluster boundary function of a meromorphic function

Let $f(z)$ be a function meromorphic for $|z|<1$, with cluster boundary function $\mathcal{F}(z)$.† Let P be a point on $|z|=1$. What are the relations between $f(z)$ and $\mathcal{F}(z)$ in a neighborhood of P? A partial answer to this has been given in §5, since if $f(z)$ is bounded, the value of its Fatou boundary function at a

* A function $f(z)$ is called univalent if $f(z_1)\neq f(z_2)$ unless $z_1=z_2$.

† Cf. §2.

point P on $|z|=1$ where it is defined is also one of the values of $\mathcal{J}(P)$. It will be seen that the results of this section are generalizations of the following theorem.

THEOREM 6.1. *Let $f(z)$ be meromorphic for $|z|<1$, and let S be the sum of the cluster sets of $f(z)$ in $|z|<1$ at all the points of $|z|=1$. If $f(z)$ takes on a value α not in S, $f(z)$ assumes every value in the domain containing α and bounded only by points of S.**

Let $f_n(z)=f(z)$, $n=1, 2, \cdots$. Then this theorem is simply Theorem 4.1 for the sequence $\{f_n(z)\}$. A simple direct proof is the following. The set of values s assumed by $f(z)$ in $|z|<1$ is open. If s does not contain the domain D considered, there is a frontier point P of s in D. The point P is a limit point of assumed points:

$$\lim_{n\to\infty} f(z_n) = P$$

for some sequence $\{z_n\}$ in $|z|<1$. Since P is not assumed in $|z|<1$,

$$\lim_{n\to\infty}|z_n| = 1,$$

and so P belongs to S, contrary to the hypothesis that P was in D.

The theorem corresponding to Theorem 4.2 in this development is an important theorem first proved by W. Gross and F. Iversen.† This theorem can be proved most easily directly,‡ although the greater part can be proved without difficulty by the methods of this paper. The following is a generalization of the Gross-Iversen theorem.

THEOREM 6.2. *Let $f(z)$ be meromorphic in γ: $\mathfrak{I}(z)>0$, and let the cluster set of $f(z)$ on the real axis at $P: z=0$ on a given point set E be denoted by $S(E)$. Let α be a point of the cluster set s of $f(z)$ in γ at P:*

$$\lim_{n\to\infty} z_n = 0, \quad \lim_{n\to\infty} f(z_n) = \alpha, \tag{6.21}$$

and let $z_n = x_n + iy_n$ where x_n and y_n are real.

---

* This theorem, a generalization of a well known theorem of Darboux, was obtained recently by Persidskij, Bulletin de la Société Physico-Mathématique de Kazan, vol. 3, No. 4 (1931), pp. 89–91.

† W. Gross, Monatshefte für Mathematik und Physik, vol. 29 (1918), pp. 3–47; Mathematische Zeitschrift, vol. 2 (1918), pp. 242–294.

F. Iversen, *Recherches sur les Fonctions Inverses des Fonctions Méromorphes*, Thesis, Helsingfors, 1914; Översikt av Finska Vetenskaps-Societetens Förhandlingar, vol. 58 (1915–1916), Section A, No. 25; ibid., vol. 64 (1921–1922), Section A, No. 4.

‡ Cf. a paper by the author in the Annals of Mathematics, (2), vol. 33 (1932), pp. 753–757.

(a) *Suppose that the sequence* $\{z_n\}$ *is tangential*:

$$\lim_{n\to\infty} \frac{y_n}{x_n} = 0.$$

Let R_1 *be the interior of a rectangle one of whose sides is on the real axis, and whose diagonals intersect at* z_1. *Let* R_n, $n>1$, *be the interior of a rectangle, one of whose sides is on the real axis and whose diagonals, parallel to those of* R_1, *intersect at* $x_n+i\eta_n$ *where* η_n *is chosen so that*

$$\eta_n > 0, \lim_{n\to\infty} \frac{\eta_n}{x_n} = 0, \quad \lim_{n\to\infty} \frac{y_n}{\eta_n} = 0.$$

Let E_N *be the set of all those points on a base of* R_n *for some value of* $n \geqq N$. *Then if* α *does not belong to* S: *the product of all the sets* $\{S(E_n)\}$, $f(z)$ *assumes in* R_n *each value in the domain* D *containing* α *and bounded only by points of* S, *for all except perhaps a finite set of values of* n, *with two possible exceptions; if there are two exceptions, they are the only ones in the extended plane for* $f(z)$ *in the rectangles* $\{R_n\}$.

(b) *Let* R_1 *be the interior of a rectangle one of whose sides is on the real axis and whose diagonals intersect at* z_1. *Let* R_n, $n>1$, *be the interior of the rectangle one of whose sides is on the real axis and whose diagonals, parallel to those of* R_1, *intersect at* z_n. *Suppose that* α *is omitted by* $f(z)$ *in the rectangles* $\{R_n\}$. *Let* E_N *be the set of all those points on a base of* R_n *for some value of* $n \geqq N$. *Then if* α *does not belong to* S: *the product of all the sets* $\{S(E_n)\}$, $f(z)$ *assumes in* R_n *each value in the domain* D *containing* α *and bounded only by points of* S *for all except perhaps a finite set of values of* n, *with one possible exception, besides* α. *If there is one other such exceptional value, it is the only other one in the extended plane for* $f(z)$ *in the rectangles* $\{R_n\}$.

In (a) the sets $S(E_n)$ are identical for large values of n. If the bases of the rectangles $\{R_n\}$ do not cover the origin in (b), $S(E_n)$ can be used in place of S, for any value of n. The results (a) and (b) are consequences of Theorem 4.2 (a) and (c) respectively, applied to $f(z)$ defined in the rectangles described, and their proof presents no difficulty.

Columbia University,
New York, N. Y.

ON FINITE-ROWED SYSTEMS OF LINEAR INEQUALITIES IN INFINITELY MANY VARIABLES. II*

BY

I. J. SCHOENBERG

1. **Introduction.** In a previous paper on the same subject† a certain class of systems of linear inequalities in infinitely many variables were solved and applications to the theory of completely monotonic functions were derived from a particular type of such systems which were called Hausdorff systems. The present paper gives an extension of those results to a similar class of systems of linear inequalities involving a double sequence of variables (§2). This extension has already been performed by T. H. Hildebrandt and myself in the very particular case of completely monotonic double sequences.‡

In §4 Hausdorff systems involving a double sequence of variables are solved and applied to extend the results of F. Hausdorff, S. Bernstein, and D. V. Widder (see I, §11) to completely monotonic functions of two variables. The results of §3 (minimal solutions, minimal representations of solutions) though not absolutely necessary for the applications made in §5, help to present them in a more elegant manner.

2. **On a certain class of systems of linear inequalities for a double sequence of variables.** Let

$$(2.1)\qquad A = \left\| \begin{matrix} a_{01} & a_{02} & a_{03} \cdots \\ a_{11} & a_{12} & a_{13} \cdots \\ a_{21} & a_{22} & a_{23} \cdots \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{matrix} \right\|, \qquad B = \left\| \begin{matrix} b_{01} & b_{02} & b_{03} \cdots \\ b_{11} & b_{12} & b_{13} \cdots \\ b_{21} & b_{22} & b_{23} \cdots \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{matrix} \right\|$$

be two given infinite matrices of real numbers. Let

$$(i_1, i_2, \cdots, i_j) = | a_{i_\alpha, \beta} |, \; [i_1, i_2, \cdots, i_j] = | b_{i_\alpha, \beta} | \quad (\alpha, \beta = 1, 2, \cdots, j).$$

Throughout this paper we shall suppose that

---

* Presented to the Society, August 31, 1932; received by the editors August 2, 1932.

† I. J. Schoenberg, *On finite-rowed systems of linear inequalities in infinitely many variables*, these Transactions, vol. 34, pp. 594–619. In the text this paper will be designated by the symbol I.

‡ See Theorem 1 of T. H. Hildebrandt and I. J. Schoenberg, *On linear functional operations and the moment problem for a finite interval in one or several dimensions*, to appear in the Annals of Mathematics. We shall frequently use the results of §3 of this paper and refer to it by the symbol HS.

$$(2.2)\qquad (i_1, i_2, \cdots, i_j) > 0, \ [i_1, i_2, \cdots, i_j] > 0 \qquad (0 \leqq i_1 < i_2 < \cdots < i_j;\ j = 1, 2, 3, \cdots).$$

Let

$$(2.3)\qquad D_1^k \xi_m \equiv \begin{vmatrix} \xi_m & a_{m,1} & \cdots & a_{m,k} \\ \xi_{m+1} & a_{m+1,1} & \cdots & a_{m+1,k} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \xi_{m+k} & a_{m+k,1} & \cdots & a_{m+k,k} \end{vmatrix},$$

$$D_2^h \eta_n \equiv \begin{vmatrix} \eta_n & b_{n,1} & \cdots & b_{n,h} \\ \eta_{n+1} & b_{n+1,1} & \cdots & b_{n+1,h} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \eta_{n+h} & b_{n+h,1} & \cdots & b_{n+h,h} \end{vmatrix} \qquad (k, h, m, n = 0, 1, 2, \cdots)$$

where $D_1^0\xi_m = \xi_m$ and $D_2^0\eta_n = \eta_n$.

Let $\mu_{mn}(m, n = 0, 1, 2, \cdots)$ be a double sequence of real variables. Both expressions

$$D_1^k (D_2^h \mu_{mn}), \ D_2^h (D_1^k \mu_{mn}),$$

in which the operator D_1^k applies to the subscript m and the operator D_2^h applies to the subscript n, are linear homogeneous combinations of the $(k+1)(h+1)$ variables $\mu_{m+k',n+h'}$ $(k'=0, 1, \cdots, k;\ h'=0, 1, \cdots, h)$. Writing out these expressions it is readily found that the coefficients of $\mu_{m+k',n+h'}$ in both expressions are equal to the product of the same two cofactors:

$$\frac{\partial}{\partial \xi_{m+k'}} D_1^k \xi_m \times \frac{\partial}{\partial \eta_{n+h'}} D_2^h \eta_n.$$

Hence

$$D_1^k (D_2^h \mu_{mn}) = D_2^h (D_1^k \mu_{mn}) = D_1^k D_2^h \mu_{mn}.$$

We shall be concerned with the problem of solving the system of linear inequalities

$$(2.4)\qquad D_1^k D_2^h \mu_{mn} \geqq 0 \qquad (k, h, m, n = 0, 1, 2, \cdots).$$

Without essentially restricting this problem, we shall suppose for convenience that

$$a_{i1} = b_{i1} = 1 \qquad (i = 0, 1, 2, \cdots).$$

As in I, Part II, and in HS, §4, we shall first solve the finite system

$$(2.5)\qquad D_1^k D_2^h \mu_{mn} \geqq 0 \qquad (k + m \leqq p;\ h + n \leqq p)$$

involving the $(p+1)^2$ variables μ_{mn} $(m, n=0, 1, \cdots, p)$. From the two identities (see I, p. 602)

$$D_1^k D_2^h \mu_{mn} = \frac{(m+1, \cdots, m+k)}{(m+1, \cdots, m+k+1)} D_1^{k+1} D_2^h \mu_{mn} + \frac{(m, \cdots, m+k)}{(m+1, \cdots, m+k+1)} D_1^k D_2^h \mu_{m+1,n},$$

$$D_1^k D_2^h \mu_{mn} = \frac{[n+1, \cdots, n+h]}{[n+1, \cdots, n+h+1]} D_1^k D_2^{h+1} \mu_{mn} + \frac{[n, \cdots, n+h]}{[n+1, \cdots, n+h+1]} D_1^k D_2^h \mu_{m,n+1}$$

and (2.2) it follows that (2.5) is equivalent to its partial system

$$D_1^{p-m} D_2^{p-n} \mu_{mn} \geqq 0 \qquad (m, n = 0, 1, \cdots, p), \tag{2.6}$$

which we shall now solve.

Let us define two new operators

$$O_1^k \bar{\xi}_m = \sum_{r=0}^{m+k} \frac{(m, r+1, \cdots, m+k)}{(r+1, \cdots, m+k)(r, \cdots, m+k)} \bar{\xi}_r,$$

$$O_2^h \bar{\eta}_n = \sum_{s=0}^{n+h} \frac{[n, s+1, \cdots, n+h]}{[s+1, \cdots, n+h][s, \cdots, n+h]} \bar{\eta}_s.$$

From the fact that the two linear transformations I(7.4) and I(7.6) are inverse to each other, it follows that the two linear transformations

$$\bar{\xi}_m = D_1^{p-m} \xi_m, \quad \xi_m = O_1^{p-m} \bar{\xi}_m \qquad (m = 0, 1, \cdots, p) \tag{2.7}$$

are inverse to each other and also that the same thing is true for the transformations

$$\bar{\eta}_n = D_2^{p-n} \eta_n, \quad \eta_n = O_2^{p-n} \bar{\eta}_n \qquad (n = 0, 1, \cdots, p). \tag{2.8}$$

Let us now consider the linear transformation

$$\rho_{pmn} = D_1^{p-m} D_2^{p-n} \mu_{mn} \qquad (m, n = 0, 1, \cdots, p). \tag{2.9}$$

From (2.7) and (2.8) we derive successively

$$D_2^{p-n} \mu_{mn} = O_1^{p-m} \rho_{pmn} \text{ and } \mu_{mn} = O_2^{p-n} O_1^{p-m} \rho_{pmn}.$$

Hence

$$\mu_{mn} = O_1^{p-m} O_2^{p-n} \rho_{pmn} \qquad (m, n = 0, 1, \cdots, p) \tag{2.10}$$

is the linear transformation inverse to (2.9): *The system* (2.10) *gives, for* $\rho_{pmn} \geqq 0$, *the most general solution of* (2.6) *and hence of* (2.5).

The explicit form of (2.10) is

$$\mu_{mn} = \sum_{r,s=0}^{p} \frac{(m, r+1, \cdots, p)}{(r, \cdots, p)(r+1, \cdots, p)} \cdot \frac{[n, s+1, \cdots, p]}{[s, \cdots, p][s+1, \cdots, p]} \rho_{prs}.$$

Introducing the quantities

$$(2.11)\quad \lambda_{prs} = \frac{(0, r+1, \cdots, p)}{(r, \cdots, p)(r+1, \cdots, p)} \cdot \frac{[0, s+1, \cdots, p]}{[s, \cdots, p][s+1, \cdots, p]} \rho_{prs}$$

we get

$$\mu_{mn} = \sum_{r,s=0}^{p} \frac{(m, r+1, \cdots, p)}{(0, r+1, \cdots, p)} \cdot \frac{[n, s+1, \cdots, p]}{[0, s+1, \cdots, p]} \lambda_{prs}$$

which we write

$$(2.12)\qquad \mu_{mn} = \sum_{r,s=0}^{p} c_{mrp} d_{nsp} \lambda_{prs} \qquad (m, n = 0, 1, \cdots, p),$$

where

$$(2.13)\qquad c_{mrp} = \frac{(m, r+1, \cdots, p)}{(0, r+1, \cdots, p)}, \quad d_{nsp} = \frac{[n, s+1, \cdots, p]}{[0, s+1, \cdots, p]}$$

$$(m, r, n, s = 0, 1, \cdots, p;\ c_{mpp} = (m)/(0) = 1;\ d_{npp} = [n]/[0] = 1).$$

Let

$$(2.14)\qquad x_{pr} = c_{1rp},\ y_{ps} = d_{1sp} \qquad (r, s = 0, 1, \cdots, p).$$

As in I, §8, let $u = P_m^{(p)}(x)$ $(p \geqq m)$ denote the polygonal line in the plane (x, u), joining the points (x_{pr}, c_{mrp}) $(r = 0, 1, \cdots, p)$; similarly let $v = Q_n^{(p)}(y)$ $(p \geqq n)$ be the polygonal line in the plane (y, v), joining the points (y_{ps}, d_{nsp}) $(s = 0, 1, \cdots, p)$.

It has been proved in I, §8, that

$$(2.15)\quad \lim_{p\to\infty} P_m^{(p)}(x) = \phi_m(x), \quad \lim_{p\to\infty} Q_n^{(p)}(y) = \psi_n(y) \qquad (m, n = 0, 1, 2, \cdots)$$

hold uniformly in x and y respectively, in the interval (0, 1); moreover $\phi_m(x)$ and $\psi_n(y)$ are the sequences of functions associated with the matrices A and B by Theorem 8.1 of I. It is shown there that the $\phi_m(x)$ are continuous, non-decreasing, convex, and

$$\phi_0(x) = 1,\ \phi_1(x) = x,\ \phi_{n+1}(0) = 0,\ \phi_n(1) = 1\ (0 \leqq x \leqq 1;\ n = 0, 1, 2, \cdots).$$

The $\psi_n(y)$ have the same properties.

From Theorem 8.1 of I we know that the most general solutions of the two systems of linear inequalities

$$(2.16)\quad D_1^k \xi_m \geqq 0 \quad (k, m = 0, 1, 2, \cdots), \quad D_2^h \eta_n \geqq 0 \quad (h, n = 0, 1, 2, \cdots)$$

are given respectively by

$$(2.17)\quad \begin{aligned} \xi_m &= \int_0^1 \phi_m(x) d\chi_1(x) \qquad (m = 0, 1, 2, \cdots), \\ \eta_n &= \int_0^1 \psi_n(y) d\chi_2(y) \qquad (n = 0, 1, 2, \cdots), \end{aligned}$$

where $\chi_1(x)$ and $\chi_2(y)$ are monotonic in $(0, 1)$. The same theorem says that the monotonic function $\chi_1(x)$ is essentially uniquely defined by the first set of equations (2.17) if and only if every function $f(x)$ which is continuous on $(0, 1)$ can be uniformly approximated as close as we want by linear combinations of functions of the sequence $\{\phi_m(x)\}$. A sequence of continuous functions $\{\phi_m(x)\}$ with this property shall be called *a base of continuous functions on* $(0, 1)$. The same definition will be used for functions of several variables.

For convenience we introduce the following

DEFINITION 2.1. *The first system* (2.16) *shall be called a determining system if and only if the corresponding sequence* $\{\phi_m(x)\}$ *is a base of continuous functions on* $(0, 1)$. *Otherwise it shall be called a non-determining system.*

The following theorem is readily proved.

THEOREM 2.1. (1) *If the solutions of the systems* (2.16), *as given by Theorem* 8.1 *of* I, *are* (2.17), *then the most general solution of the system*

$$(2.18)\quad D_1^k D_2^h \mu_{mn} \geqq 0 \qquad (k, h, m, n = 0, 1, 2, \cdots)$$

may be expressed in the form

$$(2.19)\quad \mu_{mn} = \int_0^1 \int_0^1 \phi_m(x)\psi_n(y) d_x d_y \chi(x, y) \quad (m, n = 0, 1, 2, \cdots)$$

with $\chi(x, y)$ *monotonic in the sense of Hardy and Krause* (see HS, §3), *and conversely,* (2.19) *always represents a solution of* (2.18).

(2) *A necessary and sufficient condition that the function* $\chi(x, y)$ *be uniquely defined by the set* (2.19) *and the additional conditions*

$$(2.20)\quad \begin{aligned} &\chi(0, 0) = \chi(x, 0) = \chi(0, y) = 0, \ \chi(x, y) = \chi(x + 0, y + 0), \\ &\qquad\qquad \textit{for } 0 < x \leqq 1, \ 0 < y \leqq 1, \end{aligned}$$

is that both systems (2.16) *shall be determining systems, in which case also* (2.18) *shall be called a determining system.*

Let $\mu_{mn}(m, n=0, 1, 2, \cdots)$ be a solution of (2.18). Then (2.5) holds for every value of p and therefore also all the consequences derived therefrom.

Let us define in the unit-square U $(0\leqq x\leqq 1,\ 0\leqq y\leqq 1)$ a step-function $\chi_p(x, y)$ as follows:

(a) $$\chi_p(x, 0) = \chi_p(0, y) = 0;$$

(b) $$\chi_p(x_{pr}+0, y_{ps}+0) - \chi_p(x_{pr}+0, y_{ps}-0) - \chi_p(x_{pr}-0, y_{ps}+0) + \chi_p(x_{pr}-0, y_{ps}-0) = \lambda_{prs} \quad (r, s = 0, 1, \cdots, p);^*$$

(c) $\chi_p(x, y)$ is constant in each of the rectangles

$$x_{pr} < x < x_{p,r+1},\ y_{ps} < y < y_{p,s+1},$$

and also on each of the line segments

$$x_{pr} < x < x_{p,r+1},\ y = 1;\ x = 1,\ y_{ps} < y < y_{p,s+1};$$

moreover $\chi_p(x, y) = \chi_p(x+0, y+0)$ for $0<x\leqq 1,\ 0<y\leqq 1$.

From (2.11), (2.9), (2.6) and (2.12) (for $m=n=0$) we conclude that

$$\lambda_{prs} \geqq 0 \quad (r, s = 0, 1, \cdots, p), \quad \sum_{r,s=0}^{p}\lambda_{prs} = \mu_{00}, \tag{2.21}$$

which shows that $\{\chi_p(x, y)\}$ is a sequence of uniformly bounded monotonic functions in U. On the other hand, from (2.12) we derive for $p\geqq \max(m, n)$

$$\begin{aligned}\mu_{mn} &= \sum_{r,s=0}^{p} P_m^{(p)}(x_{pr})Q_n^{(p)}(y_{ps})\lambda_{prs} = \int_0^1\int_0^1 P_m^{(p)}(x)Q_n^{(p)}(y)d_xd_y\chi_p(x, y) \\ &= \int_0^1\int_0^1 \phi_m(x)\psi_n(y)d_xd_y\chi_p(x, y) + \epsilon_{pmn},\end{aligned}$$

where $\epsilon_{pmn}\to 0$ as $p\to\infty$, on account of the uniform convergence in (2.15) and the uniform boundedness of the $\chi_p(x, y)$. From a theorem† of J. Radon, we know that there is a subsequence χ_q of χ_p converging in U to a monotonic function χ. From the same lemma and our last relation we derive (2.19) as $p=q\to\infty$.

---

* $\chi_p(1+0, y_{ps}\pm 0)$ means $\chi_p(1, y_{ps}\pm 0)$; $\chi_p(0-0, y_{ps}\pm 0)$ means $\chi_p(0, y_{ps}\pm 0)$, etc.

† This theorem, which is an extension to functions of two variables of a well known theorem of Helly, says: If $\{\chi_p(x, y)\}$ is a sequence of functions which are uniformly bounded and uniformly of bounded variation in U, then there is a subsequence $\{\chi_q(x, y)\}$ converging everywhere in U to a function $\chi(x, y)$ of bounded variation in U. Moreover, for every $f(x, y)$ continuous in U

$$\lim_{q\to\infty}\int_0^1\int_0^1 f(x, y)d_xd_y\chi_q(x, y) = \int_0^1\int_0^1 f(x, y)d_xd_y\chi(x, y).$$

J. Radon, Sitzungsberichte der Wiener Akademie, vol. 122 IIa (1913), pp. 1337–1342, and vol. 128 IIa (1919), pp. 1092–1094, proved this theorem in a slightly weaker form, which, however, would also suffice for our purpose. For the present statement see HS, §3, Lemma 1.

Conversely, let χ in (2.19) be a monotonic function. We have to show that (2.19) represents a solution of (2.18). Indeed

$$D_1^k D_2^h \mu_{mn} = \int_0^1 \int_0^1 D_1^k D_2^h \phi_m(x)\psi_n(y) d_x d_y \chi(x, y)$$

$$= \int_0^1 \int_0^1 [D_1^k \phi_m(x)][D_2^h \psi_n(y)] d_x d_y \chi(x, y) \geqq 0,$$

since $D_1^k \phi_m(x) \geqq 0$, $D_2^h \psi_n(y) \geqq 0$, for any x and y on (0, 1).

The second part of Theorem 2.1 follows readily from an extension to two variables of a theorem of F. Riesz.* From this theorem it follows that $\chi(x, y)$ is uniquely defined by (2.19) and (2.20) if and only if $\{\phi_m(x)\psi_n(y)\}$ is a base of continuous functions in U. However, it is readily shown that $\{\phi_m(x)\psi_n(y)\}$ is a base of continuous functions in U if and only if both sequences $\{\phi_m(x)\}$ and $\{\psi_n(y)\}$ are such bases in (0, 1). For if both sequences $\{\phi_m(x)\}$, $\{\psi_n(y)\}$ are bases, then every polynomial $P(x, y)$ can be uniformly approximated by expressions of the form $\sum_{m,n} \gamma_{mn}\phi_m(x)\psi_n(y)$, hence also any continuous $f(x, y)$. Conversely, if this is true for any $f(x, y)$, then in particular for any continuous $f(x)$ we have

$$f(x) = \sum \gamma_{mn}\phi_m(x)\psi_n(y) + \rho(x, y) \qquad (|\rho| < \epsilon)$$

throughout U. An integration over (0, 1) with respect to y shows that $\{\phi_m(x)\}$ is a base in (0, 1). This completes the proof of Theorem 2.1.

3. **Minimal solutions.** We have so far solved completely the following three systems of linear inequalities:

$$D_1^k \xi_m \geqq 0 \qquad (k, m = 0, 1, 2, \cdots), \tag{3.1}$$

$$D_2^h \eta_n \geqq 0 \qquad (h, n = 0, 1, 2, \cdots), \tag{3.2}$$

$$D_1^k D_2^h \mu_{mn} \geqq 0 \qquad (k, h, m, n = 0, 1, 2, \cdots), \tag{3.3}$$

and their most general solutions were found to be

* See F. Riesz, loc. cit. in I, §1. The extended theorem says: If $\phi_{mn}(x, y)$ $(m, n = 0, 1, 2, \cdots)$ is a double sequence of continuous functions in U, then a function $\chi(x, y)$ monotonic in U is uniquely defined by the set of equations

$$\mu_{mn} = \int_0^1 \int_0^1 \phi_{mn}(x, y) d_x d_y \chi(x, y) \qquad (m, n = 0, 1, 2, \cdots)$$

and the conditions (2.20), if and only if $\{\phi_{mn}(x, y)\}$ is a base of continuous functions in U. There seems to be no proof of this theorem in the literature. However, the proof for the case of one variable given by W. Seidel, Annals of Mathematics, (2), vol. 32 (1931), pp. 777–784, can be extended immediately to prove the theorem just stated, if one applies Lemma 1 of HS, §3. The same theorem may also be derived from general results concerning linear metric spaces. See I. J. Schoenberg and W. Seidel, *On linear operations in linear metric spaces*, to appear in these Transactions.

$$\xi_m = \int_0^1 \phi_m(x) d\chi_1(x) \qquad (m = 0, 1, 2, \cdots) \tag{3.1'}$$

$$(\chi_1(0) = 0,\ \chi_1(x) = \chi_1(x+0) \text{ for } 0 < x < 1),$$

$$\eta_n = \int_0^1 \psi_n(y) d\chi_2(y) \qquad (n = 0, 1, 2, \cdots) \tag{3.2'}$$

$$(\chi_2(0) = 0,\ \chi_2(y) = \chi_2(y+0) \text{ for } 0 < y < 1),$$

$$\mu_{mn} = \int_0^1 \int_0^1 \phi_m(x)\psi_n(y) d_x d_y \chi(x, y) \quad (m, n = 0, 1, 2, \cdots) \tag{3.3'}$$

$$(\chi(x, y) \text{ satisfies } (2.20)),$$

respectively, where $\chi_1(x)$, $\chi_2(y)$ and $\chi(x, y)$ are monotonic.

If ξ_m is a solution of (3.1), then also $\xi_m + (0)^m\gamma$ $(\gamma > 0)$ is such a solution.* We shall need the following

Definition 3.1. *A solution ξ_m of* (3.1) *is called a minimal solution*† *if there is no other solution ξ_m' of* (3.1) *and a constant $\gamma > 0$ such that*

$$\xi_m = \xi_m' + (0)^m\gamma \qquad (m = 0, 1, 2, \cdots).$$

We prove now

Theorem 3.1. *Let* (3.1) *be a determining system.*‡ *Its solution ξ_m given by* (3.1′) *is a minimal solution if and only if the monotonic function $\chi_1(x)$ is continuous at $x = 0$.*

The condition $\chi_1(0) = \chi_1(+0)$ is necessary for ξ_m to be a minimal solution of (3.1). For let us suppose that $0 = \chi_1(0) < \chi_1(+0)$, and let us define the function $\chi_{10}(x) = \chi_1(x) + (0)^x\chi_1(+0)$ which is continuous at the origin. Then

$$\xi_m = \int_0^1 \phi_m d\chi_1(x) = \int_0^1 \phi_m(x) d\chi_{10}(x) + (0)^m\chi(+0)$$

shows that ξ_m is no minimal solution of (3.1).

---

* We define $(0)^m = 0$ for $m > 0$, $(0)^0 = 1$. Note that $\phi_m(0) = (0)^m$, $\phi_n(0) = (0)^n$.

† Hausdorff has already called attention to the distinction between minimal and non-minimal completely monotonic sequences. The name *minimal* is due to D. V. Widder, these Transactions. vol. 33 (1931), p. 880.

‡ That the condition of this theorem is not always sufficient to insure that ξ_m is a *minimal* solution of a *non-determining* system is shown by the following example. Let (3.1) be a non-determining Hausdorff system of the type considered in Theorem 10.2 of I. Let $a_0 = 0$, hence $c_{00} = c_{01} = c_{02} = \cdots = 1$. Take $\chi_1(x) = x$ for $0 \leqq x \leqq c_{11}$, $\chi_1(x) = c_{11}$ for $c_{11} \leqq x \leqq 1$. This function is continuous throughout (0, 1). However, the solution given by (3.1′) is readily found to be $\xi_0 = c_{11}/2 + c_{11}/2$, $\xi_1 = c_{11}^2/2$, $\xi_2 = \xi_3 = \cdots = 0$ and this is a non-minimal solution since $c_{11} > 0$, and $c_{11}/2, c_{11}^2/2, 0, 0, 0, \cdots$ is the solution of (3.1) given by Theorem 10.2 of I, for $\gamma = \mu_0 = \mu_2 = \mu_3 = \cdots = 0$, $\mu_1 = c_{11}/2$.

To show the sufficiency of our condition let us prove that

$$\xi_m = \int_0^1 \phi_m(x)d\chi_1(x) \text{ with } \chi_1(0) = \chi_1(+0)$$

is a minimal solution. Indeed, if ξ_m were not a minimal solution, then we should have

$$\int_0^1 \phi_m d\chi_1(x) = \int_0^1 \phi_m d\bar{\chi}_1(x) + (0)^m\gamma$$

with $\bar{\chi}_1$ monotonic, $\bar{\chi}_1(0)=0$, and $\gamma>0$. The function $\bar{\chi}_{11}(x)=\chi_1(x)+[1-(0)^x]\gamma$ is monotonic and

$$\int_0^1 \phi_m d\chi_1(x) = \int_0^1 \phi_m d\bar{\chi}_{11}(x) \qquad (m = 0, 1, 2, \cdots)$$

which is impossible, since (3.1) is a determining system, while

$$0 = \chi_1(0) = \bar{\chi}_{11}(0) = \chi_1(+0) < \bar{\chi}_{11}(+0) = \bar{\chi}_1(+0) + \gamma.$$

For a more thorough investigation of the solutions of the system (3.3) we shall need the following

LEMMA 3.1. *Let $\chi(x, y)$ be monotonic in U and satisfy the conditions* (2.20). *If we define a new function $\chi_0(x, y)$ as follows:*

$$\begin{aligned}&\chi_0(x, y) = \chi(x, y) \textit{ for } 0 < x \leqq 1, 0 < y \leqq 1;\ \chi_0(0, 0) = \chi(+0, +0);\\ &\chi_0(x, 0) = \chi(x, +0) \textit{ for } 0 < x \leqq 1;\ \chi_0(0, y) = \chi(+0, y) \textit{ for } 0 < x \leqq 1;\end{aligned} \tag{3.4}$$

then the solution (3.3′) *of* (3.3) *may be written in the form*

$$\begin{aligned}\mu_{mn} &= \int_0^1\int_0^1 \phi_m(x)\psi_n(y)d_xd_y\chi(x, y)\\ &= \int_0^1\int_0^1 \phi_m(x)\psi_n(y)d_xd_y\chi_0(x, y)\\ &\quad + (0)^n\int_0^1 \phi_m(x)d\chi_0(x, 0)\\ &\quad + (0)^m\int_0^1 \psi_n(y)d\chi_0(0, y) + (0)^{m+n}\chi_0(0, 0).\end{aligned} \tag{3.5}$$

This follows readily from the definition of the double Stieltjes integral. We know that μ_{mn} is the limit of the following expression (see HS, §3; here we take $\xi_{ij}=\xi_i$, $\eta_{ij}=\eta_j$, with $\xi_0=\eta_0=0$)

$$\sum_{i=0}^{p-1}\sum_{j=0}^{q-1}\phi_m(\xi_i)\psi_n(\eta_j)\Delta(\chi;\, x_{i+1},\, y_{j+1},\, x_i,\, y_j)^*$$

$$= \sum_{i=0}^{p-1}\sum_{j=0}^{q-1}\phi_m(\xi_i)\psi_n(\eta_j)\Delta(\chi_0;\, x_{i+1},\, y_{j+1},\, x_i,\, y_j)$$

$$+ (0)^n \sum_{i=0}^{p-1}\phi_m(\xi_i)[\chi_0(x_{i+1}, 0) - \chi_0(x_i, 0)]$$

$$+ (0)^m \sum_{j=0}^{q-1}\psi_n(\eta_j)[\chi_0(0, y_{j+1}) - \chi_0(0, y_j)] + (0)^{m+n}\chi_0(0, 0),$$

as $p\to\infty$, $q\to\infty$, and all subintervals tend to zero. The last identity follows from (3.4). Passing to the limit, it goes over into (3.5) which is thus proved.

LEMMA 3.2. *If μ_{mn} is a solution of* (3.3) *then also*

$$\bar\mu_{mn} = \mu_{mn} + (0)^n\xi_m + (0)^m\eta_n + (0)^{m+n}\gamma, \tag{3.6}$$

where ξ_m and η_n are solutions of (3.1) *and* (3.2) *respectively, and $\gamma \geqq 0$, is a solution of* (3.3).

Let

$$\xi_m = \int_0^1 \phi_m(x)d\chi_1(x) + (0)^m\gamma_1 \quad (\chi_1(0) = \chi_1(+0) = 0,\ \gamma_1 \geqq 0), \tag{3.7}$$

$$\eta_n = \int_0^1 \psi_n(y)d\chi_2(y) + (0)^n\gamma_2 \quad (\chi_2(0) = \chi_2(+0) = 0,\ \gamma_2 \geqq 0). \tag{3.8}$$

With $\chi_0(x, y)$ defined by (3.4), we define two new functions

$$\bar\chi_0(x, y) = \chi_0(x, y) + \chi_1(x) + \chi_2(y) + \gamma_1 + \gamma_2 + \gamma \tag{3.9}$$

and

$$\bar\chi(x, y) = \bar\chi_0(x, y), \quad \bar\chi(x, 0) = \bar\chi(0, y) = \bar\chi(0, 0) = 0, \tag{3.10}$$

$$\text{for } 0 < x \leqq 1,\ 0 < y \leqq 1.$$

Then $\bar\chi(x, y)$ is a monotonic function of which the corresponding function given by (3.4) is precisely $\chi_0(x, y)$. From (3.6), (3.5), (3.7) and (3.8) we then derive

---

* We write $\chi(x_{i+1}, y_{j+1}) - \chi(x_{i+1}, y_j) - \chi(x_i, y_{j+1}) + \chi(x_i, y_j) = \Delta(\chi; x_{i+1}, y_{j+1}, x_i, y_j)$.

$$\bar{\mu}_{mn} = \int_0^1 \int_0^1 \phi_m \psi_n d_x d_y \bar{\chi}_0(x, y) + (0)^n \int_0^1 \phi_m d\bar{\chi}_0(x, 0)$$

$$+ (0)^m \int_0^1 \psi_n d\bar{\chi}_0(0, y) + (0)^{m+n} \bar{\chi}_0(0, 0)$$

$$= \int_0^1 \int_0^1 \phi_m(x) \psi_n(y) d_x d_y \bar{\chi}(x, y).$$

Hence $\bar{\mu}_{mn}$ is a solution of (3.3).

Our last lemma justifies the following

DEFINITION 3.2. *A solution $\bar{\mu}_{mn}$ of* (3.3) *is called a minimal solution if there is no other solution μ_{mn} of* (3.3), *as well as two solutions ξ_m and η_n of* (3.1) *and* (3.2), *and a constant $\gamma \geqq 0$, with $\xi_0 + \eta_0 + \gamma > 0$, such that* (3.6) *shall hold for $m, n = 0, 1, 2, \cdots$.*

A first criterion for minimal solutions of (3.3) is given by

LEMMA 3.3. *A solution μ_{mn} of the determining system* (3.3) *is a minimal solution if and only if both sequences μ_{m0} and μ_{0n} are minimal solutions of the corresponding systems* (3.1) *and* (3.2), *in which case μ_{mn} is, for any fixed value of n, a minimal solution of* (3.1), *and similarly, for any fixed value of m, a minimal solution of* (3.2).

If μ_{mn} of (3.5) is a minimal solution of (3.3), then necessarily

$$\chi_0(x, 0) \equiv \chi_0(0, y) \equiv \chi_0(0, 0) = 0, \tag{3.11}$$

otherwise the representation (3.5) would contradict the assumption that μ_{mn} is minimal. We therefore have $\chi(x, y) \equiv \chi_0(x, y)$. Integrating by parts* we get

$$\mu_{m0} = \int_0^1 \int_0^1 \phi_m(x) d_x d_y \chi(x, y) = \int_0^1 \phi_m(x) d\chi(x, 1),$$

which shows that μ_{m0} is a minimal solution of the determining system (3.1) because $\chi(+0, 1) = \chi_0(0, 1) = \chi(0, 1) = 0$ (Theorem 3.1). A similar proof shows that μ_{0n} is a minimal solution of (3.2).

Suppose now that μ_{mn} is not a minimal solution of (3.3). Then

$$\mu_{mn} = \mu'_{mn} + (0)^n \xi_m + (0)^m \eta_n + (0)^{m+n} \gamma,$$

with $\xi_0 + \eta_0 + \gamma > 0$. One of the quantities $\xi_0 + \gamma$, $\eta_0 + \gamma$ is > 0. Suppose that $\eta_0 + \gamma > 0$. For $n = 0$ we derive

* See E. W. Hobson, *Functions of a Real Variable*, vol. 1, 3d edition, 1927, §448.

$$\mu_{m0} = \mu'_{m0} + \xi_m + (0)^m(\eta_0 + \gamma),$$

and hence μ_{m0} is no minimal solution of (3.1).

Suppose again that μ_{mn} is a minimal solution of (3.3). Writing

$$\bar{\chi}(x, y) = \int_0^x \int_0^y \psi_n(y) d_x d_y \chi(x, y),$$

we get

$$\mu_{mn} = \int_0^1 \int_0^1 \phi_m(x)\psi_n(y) d_x d_y \chi(x, y) = \int_0^1 \int_0^1 \phi_m(x) d_x d_y \bar{\chi}(x, y)$$

$$= \int_0^1 \phi_m(x) d_x \bar{\chi}(x, 1),$$

while

$$\bar{\chi}(x, 1) = \int_0^x \int_0^1 \psi_n(y) d_x d_y \chi(x, y) = \int_0^1 \psi_n(y) d_y \chi(x, y)$$

is obviously continuous at $x=0$. We conclude again from Theorem 3.1 that μ_{mn} is a minimal solution of (3.1), for n fixed.

From this lemma we derive the following

Theorem 3.2. *The solution* (3.3′) *of the determining system* (3.3) *is a minimal solution of this system if and only if* $\chi(x, y)$ *is continuous as a function of* (x, y) *along the two sides of the unit-square* U *which meet at the origin.*

For convenience, we denote by L those two sides of U. If μ_{mn} is a minimal solution of (3.3), then (3.11) holds and this obviously implies the continuity of $\chi(x, y)$ along L.

Conversely, if $\chi(x, y)$ is continuous along L, then

$$\mu_{m0} = \int_0^1 \int_0^1 \phi_m(x) d_x d_y \chi(x, y) = \int_0^1 \phi_m(x) d_x \chi(x, 1)$$

is a minimal solution of the determining system (3.1), since $\chi(+0, 1)=0$ (Theorem 3.1). Similarly μ_{0n} is a minimal solution of (3.2). It therefore follows from Lemma 3.3 that μ_{mn} is a minimal solution of (3.3) and the theorem is proved.

Of importance is the following

Lemma 3.4. *Every solution* μ_{mn} *of the determining system* (3.3) *may be expressed in the form*

$$\mu_{mn} = \bar{\mu}_{mn} + (0)^n\xi_m + (0)^m\eta_n + (0)^{m+n}\gamma, \tag{3.12}$$

where $\bar{\mu}_{mn}$, ξ_m, η_n *are minimal solutions of the systems* (3.3), (3.1), (3.2), *and* $\gamma \geq 0$. *This representation is unique and shall be called the minimal representation of the solution* μ_{mn} *of the system* (3.3).

Let our solution μ_{mn} be given in the form (3.5). Introducing the new function

$$\chi_{00}(x, y) = \chi_0(x, y) - \chi_0(x, 0) - \chi_0(0, y) + \chi_0(0, 0), \tag{3.13}$$

from (3.5) we derive

$$\begin{aligned}\mu_{mn} &= \int_0^1 \int_0^1 \phi_m(x)\psi_n(y)d_xd_y\chi_{00}(x, y)\\ &+ (0)^n\int_0^1 \phi_m(x)d\chi_0(x, 0) + (0)^m\int_0^1 \psi_n(y)d\chi_0(0, y) + (0)^{m+n}\chi_0(0, 0).\end{aligned} \tag{3.14}$$

This is a representation of the type (3.12). Indeed, $\chi_0(x, 0)$ and $\chi_0(0, y)$ are both continuous at the origin and $\chi_0(0, 0) \geqq 0$. Moreover, the function $\chi_{00}(x, y)$ defined by (3.13) is readily found to be continuous along L and vanishing on L. From Theorems 3.1 and 3.2 we infer the truth of our last statement.

To prove the uniqueness of (3.12), let

$$\begin{aligned}\mu_{mn} &= \bar{\mu}_{mn} + (0)^n\xi_m + (0)^m\eta_n + (0)^{m+n}\gamma\\ &= \bar{\mu}'_{mn} + (0)^n\xi'_m + (0)^m\eta'_n + (0)^{m+n}\gamma'\end{aligned} \tag{3.15}$$

be two representations of the type (3.12). For any particular value of $n > 0$, we get

$$\bar{\mu}_{mn} = \bar{\mu}'_{mn} \qquad (m = 1, 2, 3, \cdots). \tag{3.16}$$

Since $\bar{\mu}_{mn}, \bar{\mu}$ $(m = 0, 1, 2, \cdots)$ are both minimal solutions of (3.1) (Lemma 3.3), (3.16) must hold also for $m = 0$. A similar argument applied to the subscript n, shows that (3.16) holds whenever $m+n > 0$. Then necessarily $\bar{\mu}_{00} = \bar{\mu}'_{00}$, since both $\bar{\mu}_{mn}$ and $\bar{\mu}'_{mn}$ are minimal solutions of (3.3). From (3.15) we now derive

$$(0)^n\xi_m + (0)^m\eta_n + (0)^{m+n}\gamma = (0)^n\xi'_m + (0)^m\eta'_n + (0)^{m+n}\gamma',$$

from which, for $n = 0$, $m > 0$, we get $\xi_m = \xi'_m$, which holds also for $m = 0$. Similarly $\eta_n = \eta'_n$, and finally for $m = n = 0$, we obtain $\gamma = \gamma'$.

A consequence is the following

THEOREM 3.3. *Every solution μ_{mn} of the determining system* (3.3) *may be represented as follows*:

$$\begin{aligned}\mu_{mn} &= \int_0^1 \int_0^1 \phi_m(x)\psi_n(y)d_xd_y\chi_{00}(x, y)\\ &+ (0)^n\int_0^1 \phi_m(x)d\chi_1(x) + (0)^m\int_0^1 \psi_n(y)d\chi_2(y) + (0)^{m+n}\gamma\\ &\qquad (m, n = 0, 1, 2, \cdots; \gamma \geqq 0),\end{aligned} \tag{3.17}$$

where $\chi_{00}(x, y)$, $\chi_1(x)$, $\chi_2(y)$ are monotonic and satisfy the conditions

$$\chi_{00}(0, 0) = \chi_{00}(x, 0) = \chi_{00}(0, y) = 0, \chi_{00}(x, y) = \chi_{00}(x + 0, y + 0)$$

$$\text{(3.18)} \qquad \textit{for } 0 < x \leqq 1, 0 < y \leqq 1;$$

$$\chi_{00}(x, y) \textit{ is continuous along } L(0 \leqq x \leqq 1, y = 0; x = 0, 0 \leqq y \leqq 1);$$

$$\text{(3.19)} \quad \begin{aligned} &\chi_1(0) = \chi_1(+0) = 0, \chi_1(x) = \chi_1(x + 0) \textit{ for } 0 < x < 1, \\ &\chi_2(0) = \chi_2(+0) = 0, \chi_2(y) = \chi_2(y + 0) \textit{ for } 0 < y < 1. \end{aligned}$$

This is a minimal representation of the solution μ_{mn} and is unique in the sense that the three monotonic functions $\chi_{00}(x, y)$, $\chi_1(x)$, $\chi_2(y)$ and the constant γ are uniquely defined by (3.17), (3.18), *and* (3.19). *The solution μ_{mn} is minimal if and only if $\chi_1(x) \equiv \chi_2(y) \equiv \gamma = 0$.*

From the minimal representation (3.12) and Theorems 3.1 and 3.2 we immediately derive (3.17). The uniqueness of (3.17) follows from the uniqueness of a minimal representation (Lemma 3.4) and from the fact that our systems (3.1), (3.2) and (3.3) are determining systems.

From (3.3′) we derive (3.17) by means of (3.4), (3.13) and

$$\text{(3.20)} \quad \chi_1(x) = \chi_0(x, 0) - \chi_0(0, 0), \chi_2(y) = \chi_0(0, y) - \chi_0(0, 0), \gamma = \chi_0(0, 0).$$

4. **Hausdorff systems for double sequences.** The system of linear inequalities (3.3) is called a *Hausdorff system* if both systems (3.1) and (3.2) are Hausdorff systems, that is to say (see I, §9), when both matrices A and B of (2.1) are of the Vandermondean type:

$$\text{(4.1)} \qquad A = \left\| \begin{matrix} 1 & a_0 & a_0^2 & \cdots \\ 1 & a_1 & a_1^2 & \cdots \\ 1 & a_2 & a_2^2 & \cdots \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \end{matrix} \right\|, \quad B = \left\| \begin{matrix} 1 & b_0 & b_0^2 & \cdots \\ 1 & b_1 & b_1^2 & \cdots \\ 1 & b_2 & b_2^2 & \cdots \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \end{matrix} \right\|,$$

with

$$\text{(4.2)} \qquad a_0 < a_1 < a_2 < \cdots, \quad b_0 < b_1 < b_2 < \cdots.$$

We shall invariably suppose that

$$\text{(4.3)} \qquad \sum^{+\infty} \frac{1}{a_r} = +\infty, \quad \sum^{+\infty} \frac{1}{b_r} = +\infty,$$

whenever $a_r \to +\infty$, or $b_r \to +\infty$, and discuss the following three further possibilities:

$$\text{(4.4)} \qquad \lim_{r \to +\infty} a_r = +\infty, \quad \lim_{r \to +\infty} b_r = +\infty,$$

$$\text{(4.5)} \qquad \lim_{r \to +\infty} a_r = \alpha, \quad \lim_{r \to +\infty} b_r = \beta \qquad (\alpha, \beta < +\infty),$$

$$\text{(4.6)} \qquad \lim_{r \to +\infty} a_r = \alpha, \quad \lim_{r \to +\infty} b_r = +\infty \qquad (\alpha < +\infty).$$

As we know from I, §§9–10, only the case (4.4) leads to a determining system (3.3). Moreover, in this case

$$\phi_m(x) = x^{(a_m-a_0)/(a_1-a_0)}, \quad \psi_n(y) = y^{(b_n-b_0)/(b_1-b_0)}. \tag{4.7}$$

According to Theorem 3.3, a minimal solution of the system (3.3) (with (4.1), (4.2), (4.3), and (4.4)) is given by

$$\mu_{mn} = \int_0^1 \int_0^1 x^{(a_m-a_0)/(a_1-a_0)} y^{(b_n-b_0)/(b_1-b_0)} d_x d_y \chi_{00}(x, y),$$

where $\chi_{00}(x, y)$ satisfies the conditions (3.18). Writing

$$\chi_{00}(x^{a_1-a_0}, y^{b_1-b_0}) = \chi(x, y),$$

we get

$$\mu_{mn} = \int_0^1 \int_0^1 x^{a_m-a_0} y^{b_n-b_0} d_x d_y \chi(x, y) \qquad (m, n = 0, 1, 2, \cdots), \tag{4.8}$$

where $\chi(x, y)$ satisfies the conditions (3.18) and is uniquely defined by (4.8) and (3.18).

Let us define on $U-L$ the function

$$\omega(x, y) = \int_x^1 \int_y^1 x^{-a_0} y^{-b_0} d_x d_y \chi(x, y) \qquad (0 < x \leqq 1, 0 < y \leqq 1). \tag{4.9}$$

Then, since $\chi(x, y)$ is continuous along L,

$$\mu_{mn} = \lim_{\epsilon \to 0} \int_\epsilon^1 \int_\epsilon^1 x^{a_m-a_0} y^{b_n-b_0} d_x d_y \chi(x, y) = \lim_{\epsilon \to 0} \int_\epsilon^1 \int_\epsilon^1 x^{a_m} y^{b_n} d_x d_y \omega(x, y)$$

and hence

$$\mu_{mn} = \int_0^1 \int_0^1 x^{a_m} y^{b_n} d_x d_y \omega(x, y) \qquad (m, n = 0, 1, 2, \cdots), \tag{4.10}$$

where the integrals are improper and converge in the sense that $\lim_{\epsilon \to 0} \int_\epsilon^1 \int_\epsilon^1$ exists.

Conversely let $\bar{\omega}(x, y)$ be a function defined on $U-L$, with the properties

$$\begin{gathered} \bar{\omega}(x, 1) = \bar{\omega}(1, y) = 0 \text{ for } 0 < x \leqq 1, 0 < y \leqq 1; \\ \bar{\omega}(x', y) \geqq \bar{\omega}(x'', y), \bar{\omega}(x, y') \geqq \bar{\omega}(x, y''), \\ \Delta(\bar{\omega}; x'', y'', x', y') = \bar{\omega}(x'', y'') - \bar{\omega}(x'', y') - \bar{\omega}(x', y'') \\ + \bar{\omega}(x', y') \geqq 0, \\ \text{for } 0 < x' < x'' \leqq 1, 0 < y' < y'' \leqq 1, 0 < x \leqq 1, 0 < y \leqq 1, \end{gathered} \tag{4.11}$$

and such that in the sense that $\lim_{\epsilon\to 0}\int_\epsilon^1\int_\epsilon^1$ exists,

$$\mu_{mn} = \int_0^1\int_0^1 x^{a_m}y^{b_n}d_xd_y\bar{\omega}(x, y) \qquad (m, n = 0, 1, 2, \cdots). \tag{4.12}$$

Let

$$\bar{\chi}(x, y) = \int_0^x\int_0^y x^{a_0}y^{b_0}d_xd_y\bar{\omega}(x, y) \text{ in } U, \tag{4.13}$$

from which we derive

$$\bar{\omega}(x, y) = \int_x^1\int_y^1 x^{-a_0}y^{-b_0}d_xd_y\bar{\chi}(x, y) \text{ in } U - L, \tag{4.14}$$

and

$$\mu_{mn} = \int_0^1\int_0^1 x^{a_m-a_0}y^{b_n-b_0}d_xd_y\bar{\chi}(x, y) \quad (m, n = 0, 1, 2, \cdots). \tag{4.15}$$

From (4.8), (4.15), and a theorem* of C. A. Fischer it follows that

$$\chi(x, y) = \bar{\chi}(x, y)$$

in all the points of U, except possibly for a set of points lying on two denumerable sets of vertical and horizontal line segments

$$\begin{aligned} V_i:&\quad x = \xi_i,\ 0 \leqq y \leqq 1 \quad (0 < \xi_i < 1;\ i = 1, 2, 3, \cdots),\\ H_j:&\quad y = \eta_j,\ 0 \leqq x \leqq 1 \quad (0 < \eta_j < 1;\ j = 1, 2, 3, \cdots). \end{aligned} \tag{4.16}$$

From (4.9) and (4.14) it follows that

$$\omega(x, y) = \bar{\omega}(x, y)$$

in all the points of $U-L$ outside the segments (4.16).

A consequence of these results, of Lemma 3.4, and of Corollary 9.1 of I, is the following

* The theorem referred to is equivalent to the following statement: If $g_{mn}(x, y)$ is a base of continuous functions in U, and $\chi(x, y)$, $\bar{\chi}(x, y)$ are two functions of bounded variation in U (both vanishing on L), then

$$\int_0^1\int_0^1 g_{mn}(x, y)d_xd_y\chi(x, y) = \int_0^1\int_0^1 g_{mn}(x, y)d_xd_y\bar{\chi}(x, y) \qquad (m, n = 0, 1, 2, \cdots),$$

if and only if $\chi(x, y)=\bar{\chi}(x, y)$ throughout U, except possibly for a set of points contained in two sets of line segments of the type (4.16). See C. A. Fischer, Annals of Mathematics, (2), vol. 19(1917–18), pp. 39–40, and HS, §3, Lemma 2.

THEOREM 4.1. *Every solution μ_{mn} of the determining Hausdorff system* (3.3) *derived from the matrices* (4.1), *whose elements satisfy* (4.2), (4.3), *and* (4.4), *admits the following minimal representation*:

$$(4.17)\quad \begin{aligned}\mu_{mn} = \int_0^1\int_0^1 x^{a_m}y^{b_n}d_xd_y\omega(x, y) + (0)^n\int_0^1 x^{a_m}d\rho_1(x)\\ + (0)^m\int_0^1 y^{b_n}d\rho_2(y) + (0)^{m+n}\gamma \qquad (m, n = 0, 1, 2, \cdots; \gamma \geqq 0),\end{aligned}$$

where $\omega(x, y)$, defined on $U-L$, satisfies the conditions (4.11) *and*

$$(4.18)\qquad \omega(x, y) = \omega(x-0, y-0) \quad (0 < x < 1, 0 < y < 1),$$

while $\rho_1(x)$ and $\rho_2(y)$ are monotonic on $0<x\leqq 1$ and $0<y\leqq 1$, respectively, with

$$(4.19)\qquad \begin{aligned}\rho_1(1) = \rho_2(1) = 0,\ \rho_1(x) = \rho_1(x-0),\ \rho_2(y) = \rho_2(y-0)\\ \text{for } 0 < x < 1,\ 0 < y < 1.\end{aligned}$$

The integrals in (4.17) *are improper and convergent in the sense that* $\lim_{\epsilon\to 0}\int_\epsilon^1\int_\epsilon^1$ *and* $\lim_{\epsilon\to 0}\int_\epsilon^1$ *respectively exist. The functions $\omega(x, y)$, $\rho_1(x)$, $\rho_2(y)$ and the constant γ are uniquely defined by* (4.17) *and all their further properties described above.*

Conversely, the double sequence μ_{mn} given by (4.17) *is always, if $\mu_{00}<\infty$, a solution of* (3.3).

We conclude the consideration of this case with the following remark which will be useful in the next section: *The function $\omega(x, y)$ is uniquely defined by the set of equations* (4.10) *and the condition* (4.11) *and* (4.18), *even if we leave out a finite number of equations of the set* (4.10). Let the equations with $m<m'$, $n<n'$ be left out of the set (4.10). The proof of the uniqueness of $\omega(x, y)$ is exactly the same as above, with the only difference that instead of (4.13) we associate with $\omega(x, y)$ the function

$$\chi(x, y) = \int_0^x\int_0^y x^{-a_{m'}}y^{-b_{n'}}d_xd_y\omega(x, y).$$

We consider now the second assumption (4.5). We know from I, §10, that if we write

$$(4.20)\qquad x_p = \left(\frac{\alpha - a_1}{\alpha - a_0}\right)^p, \quad y_q = \left(\frac{\beta - b_1}{\beta - b_0}\right)^q \qquad (p, q = 0, 1, 2, \cdots),$$

then $u=\phi_m(x)$ is the polygonal line which joins the vertices

$$(4.20')\qquad \left(x_p, \left(\frac{\alpha - a_m}{\alpha - a_0}\right)^p\right) \qquad (p = 0, 1, 2, \cdots) \qquad \text{and } (0, (0)^m),$$

and $v=\psi_n(y)$ is the polygonal line which joins the vertices

$$\left(y_q, \left(\frac{\beta - b_n}{\beta - b_0}\right)^q\right) \qquad (q = 0, 1, 2, \cdots) \qquad \text{and } (0, (0)^n). \tag{4.20''}$$

Let A_p be the slope of $\phi_m(x)$ on the interval $x_{p+1} \leqq x \leqq x_p$, and similarly B_q the slope of $\psi_n(y)$ on the interval $y_{q+1} \leqq y \leqq y_q$.

Let

$$\bar{\mu}_{mn} = \int_0^1 \int_0^1 \phi_m(x)\psi_n(y) d_x d_y \chi_{00}(x, y) \qquad (m, n = 0, 1, 2, \cdots) \tag{4.21}$$

be a solution of (3.3), with $\chi_{00}(x, y)$ monotonic in U, satisfying the conditions (3.18). Applying integration by parts in each subrectangle, we obtain

$$\int_{x_N}^1 \int_{y_N}^1 \phi_m(x)\psi_n(y) d_x d_y \chi_{00}(x, y) = \sum_{p,q=0}^{N-1} A_p B_q \int_{x_{p+1}}^{x_p} \int_{y_{q+1}}^{y_q} \chi_{00}(x, y) dx dy$$

$$- \int_{x_N}^1 \chi_{00}(x, 1) d\phi_m(x) - \int_{y_N}^1 \chi_{00}(1, y) d\psi_n(y) + \phi_m(x_N) \int_{y_N}^1 \chi_{00}(x_N, y) d\psi_n(y)$$

$$+ \psi_n(y_N) \int_{x_N}^1 \psi_{00}(x, y_N) d\phi_m(x) + \chi_{00}(1, 1) - \phi_m(x_N)\chi_{00}(x_N, 1)$$

$$- \psi_n(y_N)\chi_{00}(1, y_N) + \phi_m(x_N)\psi_n(y_N)\chi_{00}(x_N, y_N).$$

As $N \to \infty$, this goes over into

$$\begin{aligned} \bar{\mu}_{mn} = \sum_{p,q=0}^{\infty} A_p B_q \int_{x_{p+1}}^{x_p} \int_{y_{q+1}}^{y_q} \chi_{00}(x, y) dx dy - \sum_{p=0}^{\infty} A_p \int_{x_{p+1}}^{x_p} \chi_{00}(x, 1) dx \\ - \sum_{q=0}^{\infty} B_q \int_{y_{q+1}}^{y_q} \chi_{00}(1, y) dy + \chi_{00}(1, 1). \end{aligned} \tag{4.22}$$

Let us define in U a step-function $\bar{\chi}(x, y)$ as follows:

$\bar{\chi}(x, y) = 0$ on L; $\bar{\chi}(1, 1) = \chi_{00}(1, 1)$;

$$\bar{\chi}(x, y) = \frac{1}{(x_p - x_{p+1})(y_q - y_{q+1})} \int_{x_{p+1}}^{x_p} \int_{y_{q+1}}^{y_q} \chi_{00}(x, y) dx dy$$

$$\text{for } x_{p+1} \leqq x < x_p,\ y_{q+1} \leqq y < y_q \qquad (p, q = 0, 1, 2, \cdots);$$

$$\bar{\chi}(x, 1) = \frac{1}{x_p - x_{p+1}} \int_{x_{p+1}}^{x_p} \chi_{00}(x, 1) dx \text{ for } x_{p+1} \leqq x < x_p \qquad (p = 0, 1, 2, \cdots);$$

$$\bar{\chi}(1, y) = \frac{1}{y_q - y_{q+1}} \int_{y_{q+1}}^{y_q} \chi_{00}(1, y) dy \text{ for } y_{q+1} \leqq y < y_q \qquad (q = 0, 1, 2, \cdots).$$

This step-function is immediately found to be monotonic in U and continuous along L. Moreover, from (4.22) we infer that

$$\bar{\mu}_{mn} = \int_0^1 \int_0^1 \phi_m(x)\psi_n(y) d_x d_y \bar{\chi}(x, y) \quad (m, n = 0, 1, 2, \cdots). \tag{4.23}$$

If we write

$$\bar{\lambda}_{pq} = \bar{\chi}(x_p + 0, y_q + 0) - \bar{\chi}(x_p + 0, y_q - 0) - \bar{\chi}(x_p - 0, y_q + 0) + \bar{\chi}(x_p - 0, y_q - 0),$$

then from (4.23), (4.20′), (4.20″) and the fact that $\bar{\chi}(x, y)$ is continuous along L, we derive

$$\bar{\mu}_{mn} = \sum_{p,q=0}^{\infty} \left(\frac{\alpha - a_m}{\alpha - a_0}\right)^p \left(\frac{\beta - b_n}{\beta - b_0}\right)^q \bar{\lambda}_{pq};$$

with

$$\bar{\lambda}_{pq} = (\alpha - a_0)^p (\beta - b_0)^q \lambda_{pq}$$

we finally obtain

$$\bar{\mu}_{mn} = \sum_{p=0}^{\infty} \sum_{q=0}^{\infty} \lambda_{pq} (\alpha - a_m)^p (\beta - b_n)^q \quad (m, n = 0, 1, 2, \cdots). \tag{4.24}$$

We now immediately obtain the following

THEOREM 4.2. *Every solution μ_{mn} of the non-determining Hausdorff system* (3.3), *derived from the matrices* (4.1) *whose elements satisfy the conditions* (4.2), (4.3), *and* (4.5), *admits the following minimal representation*:

$$\begin{aligned} \mu_{mn} = {} & \sum_{p=0}^{\infty} \sum_{q=0}^{\infty} \lambda_{pq} (\alpha - a_m)^p (\beta - b_n)^q + (0)^n \sum_{p=0}^{\infty} \rho_p (\alpha - a_m)^p \\ & + (0)^m \sum_{q=0}^{\infty} \sigma_q (\beta - b_n)^q + (0)^{m+n} \gamma \end{aligned} \tag{4.25}$$

$$(m, n = 0, 1, 2, \cdots; \lambda_{pq} \geqq 0, \rho_p \geqq 0, \sigma_q \geqq 0, \gamma \geqq 0).$$

The non-negative coefficients λ_{pq}, ρ_p, σ_q and γ are all uniquely defined by the set of equations (4.25).

Conversely, the double sequence μ_{mn} given by (4.25) *is always, if $\mu_{00} < \infty$, a solution of* (3.3).

Indeed, let μ_{mn} be a solution of (3.3). From (3.3′), Lemma (3.1), and (3.13) we have

$$\begin{aligned}\mu_{mn} &= \int_0^1\int_0^1 \phi_m(x)\chi_n(y)d_xd_y\chi_{00}(x, y)\\ &\quad + (0)^n\int_0^1 \phi_m(x)d\chi_0(x, 0)\\ &\quad + (0)^m\int_0^1 \psi_n(y)d\chi_0(0, y) + (0)^{m+n}\chi_{00}(0, 0),\end{aligned}\tag{4.26}$$

where, as we know, $\chi_{00}(x, y)$ is continuous along L, and $\chi_0(x, 0)$, $\chi_0(0, y)$ are continuous at the origin.

From (4.24) and Theorem 10.1 of I, we obtain

$$\begin{aligned}\bar{\mu}_{mn} &= \int_0^1\int_0^1 \phi_m\psi_n d_xd_y\chi_{00}\\ &= \sum_{p=0}^{\infty}\sum_{q=0}^{\infty}\lambda_{pq}(\alpha - a_m)^p(\beta - b_n)^q,\end{aligned}$$

$$\xi_m = \int_0^1 \phi_m(x)d\chi_0(x, 0) = \sum_{p=0}^{\infty}\rho_p(\alpha - a_m)^p,$$

$$\eta_n = \int_0^1 \psi_n(y)d\chi_0(0, y) = \sum_{q=0}^{\infty}\sigma_q(\beta - b_n)^q,$$

which are, as is easily seen, minimal solutions of (3.3), (3.1) and (3.2) respectively. With $\chi_0(0, 0)=\gamma$, (4.26) goes over into (4.25). The uniqueness of the coefficients λ_{pq}, ρ_p, σ_q follows from known properties of power series.

We pass now to the last assumption (4.6). In this case $\phi_m(x)$ is again the polygonal line joining the points (4.20′), while

$$\psi_n(y) = y^{(b_n-b_0)/(b_1-b_0)}.\tag{4.27}$$

Let again

$$\bar{\mu}_{mn} = \int_0^1\int_0^1 \phi_m(x)\psi_n(y)d_xd_y\chi_{00}(x, y) \quad (m, n = 0, 1, 2, \cdots)\tag{4.28}$$

be a solution of (3.3), where the monotonic function $\chi_{00}(x, y)$ has the properties (3.18).

From a theorem of Fréchet* we obtain

* M. Fréchet, Nouvelles Annales de Mathématiques, (4), vol. 10 (1910), p. 253.

$$\bar{\mu}_{mn} = \int_0^1 \psi_n(y) d_y \int_0^1 \phi_m(x) d_x \chi_{00}(x, y). \tag{4.29}$$

Let us define in U a new function $\bar{\chi}(x, y)$ as follows:

$$\bar{\chi}(0, y) = \chi_{00}(0, y) = 0, \quad \bar{\chi}(1, y) = \chi_{00}(1, y) \text{ for } 0 \leqq y \leqq 1;$$

$$\bar{\chi}(x, y) = \frac{1}{x_p - x_{p+1}} \int_{x_{p+1}}^{x_p} \chi_{00}(x, y) dx \text{ for } x_{p+1} \leqq x < x_p, \ 0 \leqq y \leqq 1 \qquad (p = 0, 1, 2, \cdots).$$

This function $\chi(x, y)$ is also monotonic in U, and from I, §10, we know that

$$\begin{aligned} \int_0^1 \phi_m(x) d_x \chi_{00}(x, y) &= \int_0^1 \phi_m(x) d_x \bar{\chi}(x, y) \\ &= \sum_{p=0}^{\infty} \left(\frac{\alpha - a_m}{\alpha - a_0}\right)^p \bar{\lambda}_p(y) \text{ for } 0 \leqq y \leqq 1, \end{aligned} \tag{4.30}$$

where the

$$\bar{\lambda}_p(y) = \bar{\chi}(x_p + 0, y) - \bar{\chi}(x_p - 0, y)$$

are also monotonic functions which are continuous at $y=0$.

From (4.29), (4.30), and (4.27) we derive

$$\begin{aligned} \bar{\mu}_{mn} &= \sum_{p=0}^{\infty} \left(\frac{\alpha - a_m}{\alpha - a_0}\right)^p \int_0^1 \psi_n(y) d\bar{\lambda}_p(y) \\ &= \sum_{p=0}^{\infty} \left(\frac{\alpha - a_m}{\alpha - a_0}\right)^p \int_0^1 y^{(b_n - b_0)/(b_1 - b_0)} d\bar{\lambda}_p(y), \end{aligned}$$

and if we write

$$\lambda_p(y) = (\alpha - a_0)^{-p} \bar{\lambda}_p(y^{b_1 - b_0}),$$

this becomes

$$\bar{\mu}_{mn} = \sum_{p=0}^{\infty} (\alpha - a_m)^p \int_0^1 y^{b_n - b_0} d\lambda_p(y). \tag{4.31}$$

In this last form we easily recognize that $\bar{\mu}_{mn}$ is a minimal solution of the system (3.3). Introducing the monotonic functions

$$\omega_p(y) = -\int_y^1 y^{-b_0} d\lambda_p(y) \quad (0 < y \leqq 1; p = 0, 1, 2, \cdots),$$

(4.31) becomes

$$\bar{\mu}_{mn} = \sum_{p=0}^{\infty} (\alpha - a_m)^p \int_0^1 y^{b_n} d\omega_p(y).$$

Just as above we immediately derive the following

THEOREM 4.3. *Every solution* μ_{mn} *of the non-determining Hausdorff system* (3.3), *derived from the matrices* (4.1) *whose elements satisfy the conditions* (4.2), (4.3), *and* (4.6), *admits the following minimal representation*:

$$\begin{aligned}\mu_{mn} = {} & \sum_{p=0}^{\infty}(\alpha - a_m)^p \int_0^1 y^{b_n} d\omega_p(y) \\ & + (0)^n \sum_{p=0}^{\infty}\rho_p(\alpha - a_m)^p + (0)^m \int_0^1 y^{b_n} d\sigma(y) + (0)^{m+n}\gamma \end{aligned} \tag{4.32}$$

$$(m, n = 0, 1, 2, \cdots; \rho_p \geqq 0, \gamma \geqq 0),$$

where all the functions $\omega_p(y)$ *and* $\sigma(y)$ *are monotonic for* $0 < y \leqq 1$ *and satisfy the conditions*

$$\omega_p(1) = \sigma(1) = 0,\ \omega_p(y) = \omega_p(y - 0),\ \sigma(y) = \sigma(y - 0)\ \textit{for}\ 0 < y < 1, \tag{4.33}$$

while all the integrals are improper and converge in the sense that $\lim_{\epsilon \to 0} \int_\epsilon^1$ *exists. The functions* $\omega_p(y)$, $\sigma(y)$ *and the coefficients* ρ_p *and* γ *are uniquely defined by* (4.32) *and* (4.33).

Conversely, the double sequence μ_{mn} *given by* (4.32) *and* (4.33) *is always, if* $\mu_{00} < +\infty$, *a solution of the system* (3.3).

From the results of this section we derive easily the solutions of Hausdorff systems of a somewhat different kind. Consider the two Vandermondean matrices

$$A' = \left\Vert \begin{matrix} \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \\ 1 & a_{-1} & a_{-1}^2 & \cdots \\ 1 & a_0 & a_0^2 & \cdots \\ 1 & a_1 & a_1^2 & \cdots \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \end{matrix} \right\Vert,$$

$$B' = \left\Vert \begin{matrix} \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \\ 1 & b_{-1} & b_{-1}^2 & \cdots \\ 1 & b_0 & b_0^2 & \cdots \\ 1 & b_1 & b_1^2 & \cdots \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \end{matrix} \right\Vert, \tag{4.34}$$

where a_m and b_n are two increasing sequences for $-\infty < m < +\infty$, $-\infty < n < +\infty$, and satisfying besides (4.3) one of the conditions (4.4), (4.5) or (4.6). An immediate consequence of Theorems 4.1, 4.2, and 4.3 is the following

COROLLARY 4.1. *The most general solution of the new type of Hausdorff system*

$$D_1^k D_2^h \mu_{mn} \geqq 0 \qquad (k, h = 0, 1, 2, \cdots ; m, n = 0, \pm 1, \pm 2, \cdots) \tag{4.35}$$

is given by

$$\mu_{mn} = \int_0^1 \int_0^1 x^{a_m} y^{b_n} d_x d_y \omega(x, y), \tag{4.36}$$

$$\mu_{mn} = \sum_{p=0}^{+\infty} \sum_{q=0}^{+\infty} \lambda_{pq}(\alpha - a_m)^p (\beta - b_n)^q \qquad (\lambda_{pq} \geqq 0), \tag{4.37}$$

$$\mu_{mn} = \sum_{p=0}^{+\infty} (\alpha - a_m)^p \int_0^1 y^{b_n} d\omega_p(y), \tag{4.38}$$

for $m, n=0, \pm 1, \pm 2, \cdots$, *these three representations corresponding respectively to the possible assumptions* (4.4), (4.5), *and* (4.6). *The functions* $\omega(x, y)$, $\omega_p(y)$ *and the coefficients* λ_{pq} *enjoy the properties, in particular the uniqueness properties, described in Theorems* 4.1, 4.2, *and* 4.3.

5. **Completely monotonic functions of two variables.** A function $f(x)$ was said to be completely monotonic in an open interval (see I, §11) if it possessed derivatives of every order and

$$(-1)^p f^{(p)}(x) \geqq 0 \qquad (p = 0, 1, 2, \cdots)$$

throughout this interval.

Let $f(x, y)$ be defined in an open region R. We shall say that $f(x, y)$ is *completely monotonic* in R, if all the partial derivatives of $f(x, y)$ of every order exist and

$$(-1)^{p+q} \frac{\partial^{p+q} f(x, y)}{\partial x^p \partial y^q} \geqq 0 \qquad (p, q = 0, 1, 2, \cdots) \tag{5.1}$$

throughout the region R.

Hausdorff, Bernstein, and Widder have characterized the completely monotonic functions of one variable (see I, §11). *In this section we shall determine all the functions* $f(x, y)$ *which are completely monotonic in a rectangular region*

$$R(\alpha_0, \beta_0, \alpha, \beta): \ \alpha_0 < x < \alpha, \ \beta_0 < y < \beta \tag{5.2}$$
$$(-\infty \leqq \alpha_0 < \alpha \leqq +\infty, -\infty \leqq \beta_0 < \beta \leqq +\infty).$$

All possible cases will be taken care of if we consider successively the

following three regions:

(5.2′) $\alpha_0 < x < +\infty, \quad \beta_0 < y < +\infty,$

(5.2″) $\alpha_0 < x < \alpha, \quad \beta_0 < y < \beta \qquad (\alpha, \beta < +\infty),$

(5.2‴) $\alpha_0 < x < \alpha, \quad \beta_0 < y < +\infty \qquad (\alpha < +\infty),$

where α_0 and β_0 are either finite or else $= -\infty$.

Let us first assume that $f(x, y)$ is completely monotonic in the region (5.2′). Consider the Hausdorff system (4.35) derived from the matrices (4.34) whose elements enjoy, besides (4.3) and (4.4), the further property

$$\lim_{r\to-\infty} a_r = \alpha_0, \quad \lim_{r\to-\infty} b_r = \beta_0. \tag{5.3}$$

It follows immediately from (5.1) that for any fixed value of $x(>\alpha_0)$ and for $p \geqq 0$, the function

$$g(y) = (-1)^p(\partial^p/\partial x^p)f(x, y)$$

is completely monotonic in y for $\beta_0 < y < +\infty$. Hence

$$G(x) = \overset{h}{D_2}f(x, b_n)$$

is completely monotonic for $\alpha_0 < x < +\infty$, since

$$(-1)^p G^{(p)}(x) = \overset{h}{D_2}g(b_n) \geqq 0$$

(see I, §11, formula (11.3)). For the same reason we have

$$\overset{k}{D_1}\overset{h}{D_2}f(a_m, b_n) = \overset{k}{D_1}[\overset{h}{D_2}f(a_m, b_n)] = \overset{k}{D_1}G(a_m) \geqq 0$$

and therefore

$$\mu_{mn} = f(a_m, b_n) \qquad (m, n = 0, \pm 1, \pm 2, \cdots) \tag{5.4}$$

is a solution of the Hausdorff system (4.35).

From Corollary 4.1 we then derive

$$f(a_m, b_n) = \int_0^1 \int_0^1 \xi^{a_m}\eta^{b_n} d_\xi d_\eta \omega(\xi, \eta) \qquad (m, n = 0, \pm 1, \pm 2, \cdots)$$

where $\omega(\xi, \eta)$ has the properties given by Theorem 4.1. From the remark following Theorem 4.1, we infer that any element of the sequences a_m, b_n may vary without affecting the function $\omega(\xi, \eta)$, hence

$$f(x, y) = \int_0^1 \int_0^1 \xi^x \eta^y d_\xi d_\eta \omega(\xi, \eta) \text{ in } R(\alpha_0, \beta_0, +\infty, +\infty).$$

The transformation

$$(5.5) \qquad \xi = e^{-u}, \ \eta = e^{-v}, \ \omega(e^{-u}, e^{-v}) = \tau(u, v)$$

leads to the following

THEOREM 5.1. *Every function $f(x, y)$ which is completely monotonic in the region (5.2′) admits in this region the following representation:*

$$(5.6) \qquad f(x, y) = \int_0^{+\infty} \int_0^{+\infty} e^{-xu-yv} d_u d_v \tau(u, v),$$

where the function $\tau(u, v)$ has the following properties:

$$(5.7) \qquad \begin{aligned} &\tau(0, 0) = \tau(u, 0) = \tau(0, v) = 0, \ \tau(u, v) = \tau(u + 0, v + 0),\\ &\tau(u', v) \leqq \tau(u'', v), \ \tau(u, v') \leqq \tau(u, v''),\\ &\Delta(\tau; u'', v'', u', v') \equiv \tau(u'', v'') - \tau(u'', v') - \tau(u', v'') + \tau(u', v') \geqq 0,\\ &\textit{for } 0 < u < +\infty, 0 < v < +\infty, 0 \leqq u' < u'' < +\infty, 0 \leqq v' < v'' < +\infty. \end{aligned}$$

The improper Stieltjes integral (5.6) is absolutely convergent in $R(\alpha_0, \beta_0, +\infty, +\infty)$ and the function $\tau(u, v)$ is uniquely defined by (5.6) and (5.7).

Conversely, every function $f(x, y)$ defined by (5.6) and (5.7) is always, if finite throughout R, a completely monotonic function in $R(\alpha_0, \beta_0, +\infty, +\infty)$.

The properties (5.7) as well as the uniqueness of $\tau(u, v)$ follow from the properties (4.11), (4.18), and the uniqueness of $\omega(\xi, \eta)$, by means of the transformation (5.5). The last sentence of the theorem follows from the relations

$$(-1)^{p+q} \frac{\partial^{p+q} f(x, y)}{\partial x^p \partial y^q} = \int_0^{+\infty} \int_0^{+\infty} u^p v^q e^{-xu-yv} d_u d_v \tau(u, v) \geqq 0,$$

which are a consequence of (5.6) and where the double integral converges throughout the region R.

A particular consequence of Theorem 5.1 is that $f(x, y)$ is a real analytic and regular function of the real variables x and y in $R(\alpha_0, \beta_0, +\infty, +\infty)$.

We pass now to the consideration of a function $f(x, y)$ which is completely monotonic in the region (5.2″). Let the Hausdorff system (4.35) be defined by two sequences a_m, b_n with the properties (4.3), (4.5), and (5.3). Just as in the previous case, we conclude that (5.4) is a solution of the system (4.35). Hence from Corollary 4.1 we derive

$$f(a_m, b_n) = \sum_{p=0}^{+\infty} \sum_{q=0}^{+\infty} \lambda_{pq} (\alpha - a_m)^p (\beta - b_n)^q \quad (m, n = 0, \pm 1, \pm 2, \cdots)$$

and since again any of the numbers a_m or b_n may vary without affecting the coefficients λ_{pq}, we derive the following

THEOREM 5.2. *Every function $f(x, y)$ which is completely monotonic in the region* (5.2″), *admits in this region the following representation*:

$$f(x, y) = \sum_{p=0}^{+\infty} \sum_{q=0}^{+\infty} \lambda_{pq}(\alpha - x)^p(\beta - y)^q, \tag{5.8}$$

where $\lambda_{pq} \geqq 0$, from which it follows that $f(x, y)$ may be analytically extended and is still represented by (5.8) *in the region*

$$\alpha_0 < x < 2\alpha - \alpha_0, \quad \beta_0 < y < 2\beta - \beta_0. \tag{5.9}$$

Conversely, every function $f(x, y)$ defined by (5.8), *with $\lambda_{pq} \geqq 0$, is always, if finite throughout R, a completely monotonic function in $R(\alpha_0, \beta_0, \alpha, \beta)$.*

The last remark follows from the relations

$$(-1)^{p+q}\frac{\partial^{p+q}f(x, y)}{\partial x^p \partial y^q} = p!q! \sum_{p'=p}^{+\infty} \sum_{q'=q}^{+\infty} \binom{p'}{p}\binom{q'}{q}\lambda_{p'q'}(\alpha - x)^{p'-p}(\beta - y)^{q'-q} \geqq 0,$$

which follows from (5.8) throughout $R(\alpha_0, \beta_0, \alpha, \beta)$.

Let us finally consider a function $f(x, y)$ which is completely monotonic in the region (5.2‴). Let the Hausdorff system (4.35) be defined by two sequences a_m, b_n with the properties (4.3), (4.6), and (5.3). As in the two previous cases we derive from Corollary 4.1 the representation

$$f(x, y) = \sum_{p=0}^{\infty}(\alpha - x)^p \int_0^1 \eta^y d\omega_p(\eta) \quad \text{in} \quad R(\alpha_0, \beta_0, \alpha, +\infty),$$

and hence the following

THEOREM 5.3. *Every function $f(x, y)$ which is completely monotonic in the region* (5.2‴) *admits in this region the following representation*:

$$f(x, y) = \sum_{p=0}^{+\infty} g_p(y)(\alpha - x)^p, \tag{5.10}$$

where the functions $g_p(y)$ are completely monotonic for $\beta_0 < y < +\infty$. The representation (5.10), *which is also unique, converges and gives an analytic extension of $f(x, y)$ in the region*

$$\alpha_0 < x < 2\alpha - \alpha_0, \quad \beta_0 < y < +\infty. \tag{5.11}$$

Conversely, every function $f(x, y)$ defined by (5.10) *with the $g_p(y)$ completely monotonic for $\beta_0 < y < +\infty$, is always, if finite throughout R, a completely monotonic function in $R(\alpha_0, \beta_0, \alpha, +\infty)$.*

The last converse statement follows from the relations

$$(-1)^{p+q}\frac{\partial^{p+q}f(x,y)}{\partial x^p\partial y^q} = p!\sum_{p'=p}^{+\infty}(-1)^q g_{p'}^{(q)}(y)\binom{p'}{p}(\alpha-x)^{p'-p} \geqq 0,$$

which follows from (5.10) throughout the region $R(\alpha_0, \beta_0, \alpha, +\infty)$.

On account of the results of this section we may express the results of §4 in the following

COROLLARY 5.1. (1) *The most general solution of the Hausdorff system*

$$D_1^k D_2^h \mu_{mn} \geqq 0 \qquad (k, h = 0, 1, 2, \cdots; m, n = 0, \pm 1, \pm 2, \cdots), \tag{5.12}$$

defined by the matrices (4.34), *whose elements form increasing sequences satisfying the conditions* (4.3), *is given by*

$$\mu_{mn} = f(a_m, b_n) \qquad (m, n = 0, \pm 1, \pm 2, \cdots), \tag{5.13}$$

where $f(x, y)$ *is a function which is completely monotonic in the region*

$$\lim_{r\to-\infty} a_r < x < \lim_{r\to+\infty} a_r, \quad \lim_{r\to-\infty} b_r < y < \lim_{r\to+\infty} b_r. \tag{5.14}$$

The function $f(x, y)$ *is uniquely defined.*

(2) *The most general solution of the Hausdorff system*

$$D_1^k D_2^h \mu_{mn} \geqq 0 \qquad (k, h, m, n = 0, 1, 2, \cdots), \tag{5.15}$$

defined by the matrices (4.1), *whose elements satisfy the conditions* (4.2) *and* (4.3), *is given by*

$$\mu_{mn} = f(a_m, b_n) + (0)^n g_1(a_m) + (0)^m g_2(b_n) + (0)^{m+n}\gamma \qquad (\gamma \geqq 0), \tag{5.16}$$

where the functions $f(x, y)$, $g_1(x)$ *and* $g_2(y)$ *are completely monotonic for*

$$a_0 \leqq x < \lim_{r\to+\infty} a_r, \quad b_0 \leqq y < \lim_{r\to+\infty} b_r \tag{5.17}$$

and are uniquely defined by the relations (5.16).

In the second part of this theorem it is understood that $f(x, y)$ is completely monotonic in the interior of the region (5.17) and also continuous on the part of the boundary which belongs to this region.

UNIVERSITY OF CHICAGO,
CHICAGO, ILL.

SUFFICIENT CONDITIONS FOR THE GENERAL PROBLEM OF MAYER WITH VARIABLE END POINTS*

BY

M. R. HESTENES

1. Introduction. The problem of the calculus of variations to be considered here is the general problem of Mayer with variable end points as proposed by Bliss (V, p. 305)† and recently studied for a particular case in a joint paper by Bliss and Hestenes (XVI). As was remarked in the latter paper the general problem of Mayer is equivalent to the problem of Bolza, but the sets of sufficient conditions which have been given by Morse and Bliss for the problem of Bolza are not applicable to the problem of Mayer without further modification. In view of this fact it is the purpose of the present paper to establish a set of sufficient conditions for the general problem of Mayer with variable end points. The proofs here given are equally applicable to the problem of Bolza considered as a problem of Mayer.

The procedure used is similar to that used by Bliss for the problem of Bolza (XII, pp. 261–274). We first derive in §4 a further necessary condition analogous to that deduced by Bliss for the problem of Bolza. In §5 we construct an auxiliary problem of Mayer of the type discussed by Bliss and Hestenes (XVI). Their results are then applied in §§6 and 8 to the general problem of Mayer by methods closely related to those suggested by Mayer (XIII, pp. 436–465) and Hahn (XIV, pp. 127–136).

2. Statement of the problem. In the following pages the notation and the terminology used by Bliss and Hestenes for a particular problem of Mayer will be used throughout (XVI, pp. 306–309). In addition it will be understood that the indices μ, ν have the ranges

$$\mu, \nu = 1, \cdots, p < 2n+1.$$

The general problem of Mayer is then that of minimizing a function $g[x_1, y(x_1), x_2, y(x_2)]$ in a class of arcs

$$y_i = y_i(x) \qquad (x_1 \leqq x \leqq x_2) \tag{2:1}$$

which satisfy the differential equations and end conditions

$$\phi_\alpha(x, y, y') = 0, \quad \psi_\mu[x_1, y(x_1), x_2, y(x_2)] = 0.$$

* Presented to the Society, April 8, 1932; received by the editors June 9, 1932, and, revised, December 10, 1932.

† The Roman numerals in the parentheses in the text refer to the bibliographies at the end of the paper by Bliss and Hestenes, cited here as XVI, and at the end of the present paper.

As before, the arcs (2:1) and the functions ϕ_α, g, ψ_μ will be assumed to have the continuity properties (a), (b), (c) (XVI, p. 306) in a neighborhood of a particular arc E_0 whose minimizing properties are to be studied, the determinant (2:1) appearing in (c) being now interpreted as a $(2n+2)\times(p+1)$-dimensional matrix of rank $p+1$.

For the general problem of Mayer the first necessary condition as given by Bliss and Hestenes (XVI, p. 307) is modified as follows, and is readily established by the methods which they suggest. The theorem has also been established by Morse and Myers (X, p. 245).

I. THE FIRST NECESSARY CONDITION. *Every minimizing arc E_0 for the problem of Mayer with variable end points must satisfy, besides the conditions* (XVI, p. 307)

$$F_{y_i'} = \int_{x_1}^{x} F_{y_i} dx + c_i, \qquad \phi_\alpha = 0, \tag{2:2}$$

the further relation

$$(F - y_i' F_{y_i'})dx + F_{y_i'} dy_i \Big|_1^2 + \lambda_0 dg = 0 \tag{2:3}$$

for every set of differentials dx_1, dy_{i1}, dx_2, dy_{i2} satisfying the equations $d\psi_\mu = 0$, λ_0 being a suitably chosen constant.

An admissible arc E_0 is said to be *normal relative to the end conditions* $\psi_\mu = 0$ if there exist for it p sets of admissible variations ξ_1^ν, ξ_2^ν, $\eta_i^\nu(x)$ such that the determinant $|\Psi_\mu(\xi^\nu, \eta^\nu)|$ is different from zero (XVI, p. 307). For convenience an arc that is normal relative to the end conditions $\psi_\mu = 0$ will be designated simply as *normal*.

THEOREM 2:1. *An admissible arc that does not satisfy the necessary condition* I *is normal.*

This follows at once because an admissible arc E_0 satisfies the necessary condition I if and only if every determinant of the form

$$\begin{vmatrix} G(\xi^\sigma, \eta^\sigma) \\ \Psi_\mu(\xi^\sigma, \eta^\sigma) \end{vmatrix} \qquad (\sigma = 1, \cdots, p+1)$$

vanishes, where ξ_1^σ, ξ_2^σ, $\eta_i^\sigma(x)$ are $p+1$ sets of admissible variations for E_0, and the function G is obtained from g in the same manner as Ψ_μ is obtained from ψ_μ (V, p. 309).

THEOREM 2:2. *An admissible arc E_0 that satisfies the necessary condition* I *is normal if and only if there exist for it no set of multipliers $\lambda_\alpha(x)$, not vanishing*

simultaneously, with which it satisfies equations (2:2) *and for which all* $(p+1)$-*rowed determinants of the matrix*

$$\left\| \begin{matrix} -y_{i1}'F_{y_i'}(x_1) & F_{y_i'}(x_1) & y_{i2}'F_{y_i'}(x_2) & -F_{y_i'}(x_2) \\ \psi_{\mu x_1} & \psi_{\mu y_{i1}} & \psi_{\mu x_2} & \psi_{\mu y_{i2}} \end{matrix} \right\| \tag{2:4}$$

vanish. If E_0 *is normal the constant* λ_0 *can be chosen to be unity, the multipliers* $\lambda_\alpha(x)$ *with which* E_0 *satisfies the conditions* (2:2) *and* (2:3) *being then unique.*

This theorem is an obvious generalization of a theorem given by Bliss and Hestenes and can be proved by the same methods (XVI, p. 308). A similar theorem has been established by Bolza (III, p. 441).

3. **Theorems on extremals.** It is known that in the problems of Mayer a non-singular extremal arc can be imbedded in a $(2n-1)$-parameter family of extremals (XVI, p. 311)

$$y_i = y_i(x, c_1, \cdots, c_{2n-1}), \lambda_\alpha = \lambda_\alpha(x, c_1, \cdots, c_{2n-1}) \qquad (x_1 \leqq x \leqq x_2). \tag{3:1}$$

Further properties of this family are given in the following theorem:

THEOREM 3:1. *Let* E_0 *be a member of the* $(2n-1)$-*parameter family of extremals* (3:1) *for parameter values* (x_{10}, x_{20}, c_0). *If the matrix*

$$\left\| \begin{matrix} y_{ic_s}(x_1, c) \\ y_{ic_s}(x_2, c) \end{matrix} \right\| \tag{3:2}$$

has rank $2n-1$ *on* E_0, *then there is a neighborhood* N *of the ends of* E_0 *in* $(x_1y_1x_2y_2)$-*space such that the end values of every extremal of the family* (3:1) *with ends in* N *satisfy a relation* $W(x_1, y_1, x_2, y_2)=0$. *Conversely, every pair of points* (x_1, y_1), (x_2, y_2) *in* N *satisfying the condition* $W=0$ *can be joined by an extremal* E *of the family* (3:1), *and by taking* N *sufficiently small the parameters* (x_1, x_2, c) *belonging to* E *will lie in a preassigned* ϵ-*neighborhood of those belonging to* E_0. *The function* W *has continuous partial derivatives of the first two orders in* N.

The theorem can be proved as follows. Select $2n$ constants a_i, b_i such that the determinant

$$\left| \begin{matrix} y_{ic_s}(x_1, c) & a_i \\ y_{ic_s}(x_2, c) & b_i \end{matrix} \right| \tag{3:3}$$

is different from zero on E_0. Consider now the equations

$$\begin{aligned} y_{i1} &= y_i(x_1, c) + Wa_i, \\ y_{i2} &= y_i(x_2, c) + Wb_i. \end{aligned} \tag{3:4}$$

These equations are satisfied by the set $(x_{10}, y_{10}, x_{20}, y_{20}, c_0, W=0)$ belonging to E_0. Furthermore the functional determinant with respect to the variables c_s, W is the determinant (3:3) and is therefore different from zero on E_0. It

follows that equations (3:4) have a unique solution

$$c_s = c_s(x_1, y_1, x_2, y_2), \quad W = W(x_1, y_1, x_2, y_2) \tag{3:5}$$

in a neighborhood N of the end values $(x_{10}, y_{10}, x_{20}, y_{20})$ belonging to E_0. The right members of equations (3:5) have continuous first and second derivatives in N since the right and left members of equations (3:4) have such derivatives. If now the end values of an extremal are in N, then these end values must satisfy the relation $W(x_1, y_1, x_2, y_2) = 0$ since the solutions of equations (3:4) are unique. Furthermore every set of values (x_1, y_1, x_2, y_2) in N satisfying the relation $W=0$ are the end values of an extremal E with parameter values $[x_1, x_2, c(x_1, y_1, x_2, y_2)]$, and by taking N sufficiently small these parameter values will lie in a preassigned ϵ-neighborhood of those belonging to E_0. Hence the theorem is proved.

It is now possible to give an interesting geometric interpretation of normality.

THEOREM 3:2. *A non-singular extremal arc E_0 whose matrix* (3:2) *has rank $2n-1$ is normal if and only if in the space of points (x_1, y_1, x_2, y_2) the extremal end manifold $W=0$ and the terminal manifold $\psi_\mu=0$ are not tangent to each other at the point $(x_{10}, y_{10}, x_{20}, y_{20})$ defining the end values of E_0.*

To prove this it is sufficient, as is readily seen, to show that the derivatives $W_{x_1}, W_{y_{i1}}, W_{x_2}, W_{y_{i2}}$ are proportional to the elements of the first row of the matrix (2:4). These derivatives have this property because the relation $F_{y_i'}\eta_i$ $=$constant along extremals (XVI, p. 307) with $\eta_i = y_{ic_s}dc_s$ implies that the differentials $dx_1, dy_{i1}, dx_2, dy_{i2}, dc_s, dW$ belonging to equations (3:4) satisfy the relation

$$F_{y_i'}(dy_i - y_i'\,dx)\Big|_1^2 = F_{y_i'}y_{ic_s}dc_s\Big|_1^2 + h\,dW = h\,dW$$

where $h = b_iF_{y_i'}(x_2) - a_iF_{y_i'}(x_1)$. If $h=0$ on E_0 then on account of the relation $F_{y_i'}\eta_i$=constant, the determinant (3:3) would vanish on E_0 which is not the case. Hence $h \neq 0$ on E_0 and

$$\begin{aligned} y_{i1}'F_{y_i'}(x_1) &= hW_{x_1}, & -F_{y_i'}(x_1) &= hW_{y_{i1}}, \\ -y_{i2}'F_{y_i'}(x_2) &= hW_{x_2}, & F_{y_i'}(x_2) &= hW_{y_{i2}} \end{aligned} \tag{3:6}$$

as was to be proved.

4. The necessary condition of Mayer. The necessary condition of Mayer for the problem of Bolza, as stated by Bliss (XII, p. 266), is also valid for the problem of Mayer considered here.* In order to derive this condition we sup-

* The proof is somewhat different from that given by Bliss for the problem of Bolza. He has called my attention to the fact that the argument which he used is inadequate in the case when the ends of E_0 are conjugate, and has suggested the modifications indicated here.

pose that E_0 is a normal non-singular minimizing arc without corners and hence an extremal arc. If $\xi_1, \xi_2, \eta_i(x)$ are a set of admissible variations for E_0 which satisfy the conditions $\Psi_\mu(\xi, \eta)=0$, then E_0 is a member of a one-parameter family of admissible arcs with ends satisfying the conditions $\psi_\mu=0$ and having $\xi_1, \xi_2, \eta_i(x)$ as its variations along E_0 (IX, p. 695). For such a family the second variation of the function g to be minimized is expressible along E_0 in the form

$$(4:1) \qquad I_2 = (F_x + y_i' F_{y_i'})\xi^2 + 2F_{y_i'}\eta_i\xi \Big|_1^2 + 2(Q + l_\mu Q_\mu) + \int_{x_1}^{x_2} 2\omega(x, \eta, \eta')dx$$

where Q, Q_μ are quadratic forms in $\xi_1, \eta_i(x_1), \xi_2, \eta_i(x_2)$ whose coefficients are the second derivatives of the functions g, ψ_μ, respectively, and

$$2\omega(x, \eta, \eta') = F_{y_i y_k}\eta_i\eta_k + 2F_{y_i y_k'}\eta_i\eta_k' + F_{y_i' y_k'}\eta_i'\eta_k' .$$

This form for I_2 is readily obtained with the help of the transversality condition (2:3) by the methods used by Bliss and Hestenes (XVI, pp. 311–312). Let us consider variations satisfying the equations $\Psi_\mu(\xi, \eta)=0$ along E_0, and of the special form $\xi_1=dx_1$, $\xi_2=dx_2$, $\eta_i=\delta y_i=y_{ic_s}dc_s$, where the functions $y_i(x, c)$ are those defining the $(2n-1)$-parameter family (3:1) of extremals to which E_0 belongs. For such variations the second variation (4:1) can also be expressed in the form

$$(4:2) \qquad \begin{aligned} d^2g &= (F_x + y_i' F_{y_i'})dx^2 + 2F_{y_i'}\delta y_i dx \\ &\quad + \delta y_i \Omega_{\eta_i'}(x, \delta y, \delta y', \delta\lambda)\Big|_1^2 + 2(Q + l_\mu Q_\mu) \end{aligned}$$

given by Bliss (XII, p. 266), where $\delta\lambda_\alpha=\lambda_{\alpha c_s}dc_s$ and

$$\Omega(x, \eta, \eta', \mu) = \omega(x, \eta, \eta') + \mu_\alpha(\phi_{\alpha y_i}\eta_i + \phi_{\alpha y_i'}\eta_i').$$

Since E_0 is a minimizing arc the expression (4:1) must be $\geqq 0$ for all sets of admissible variations $\xi_1, \xi_2, \eta_i(x)$ which satisfy the conditions $\Psi_\mu(\xi, \eta)=0$. In particular it must be $\geqq 0$ for variations $\xi_1=dx_1$, $\xi_2=dx_2$, $\eta_i=\delta y_i$ of the special type considered above satisfying the conditions $d\psi_\mu \equiv \Psi_\mu(dx, \delta y)=0$. We have therefore the following result:

IV. THE NECESSARY CONDITION OF MAYER. *For a normal non-singular minimizing arc E_0 without corners the quadratic form (4:2) must satisfy the condition $d^2g \geqq 0$ for all sets $(dx_1, dx_2, dc_s) \neq (0, 0, 0)$ which satisfy the equations $d\psi_\mu=0$. Furthermore between the end points 1 and 2 on E_0 there can be no point 3 conjugate to 1 defined by a value x_3 such that E_0 is normal on the interval x_3x_2.*

The last statement is a slight modification of the condition IV deduced by Bliss and Hestenes for problems of Mayer having $2n+1$ end conditions (XVI, p. 315), valid here for E_0 since E_0 must also be a minimizing arc for such a problem, as will be seen in the next section.

5. An auxiliary problem of Mayer. In order to construct a problem of Mayer of the type described in the last paragraph we suppose that E_0 is a minimizing arc for the general problem of Mayer considered here. Its end values $(x_{10}, y_{10}, x_{20}, y_{20})$ satisfy the conditions $\psi_\mu=0$ $(\mu=1, \cdots, p)$. Adjoin to the functions ψ_μ, $2n+1-p$ functions $\psi_\tau(x_1, y_1, x_2, y_2)$ $(\tau=p+1, \cdots, 2n+1)$ possessing continuous first and second partial derivatives in a neighborhood of the values $(x_{10}, y_{10}, x_{20}, y_{20})$, vanishing at these values, and having the determinant

$$
(5:1) \qquad \begin{vmatrix} g_{x_1} & g_{y_{i1}} & g_{x_2} & g_{y_{i2}} \\ \psi_{\rho x_1} & \psi_{\rho y_{i1}} & \psi_{\rho x_2} & \psi_{\rho y_{i2}} \end{vmatrix}
$$

different from zero on E_0. The new set of end conditions $\psi_\rho=0$ $(\rho=1, \cdots, 2n+1)$ defines an auxiliary problem of Mayer of the type discussed by Bliss and Hestenes. It is clear that E_0 is also a minimizing arc for this auxiliary problem.

THEOREM 5:1. *Let E_0 be an admissible arc that is normal on the interval $x_{10}x_{20}$ and satisfies the necessary condition* I. *If E_0 is normal relative to the end conditions $\psi_\mu=0$ $(\mu=1, \cdots, p)$, then it is normal relative to the end conditions $\psi_\rho=0$ $(\rho=1, \cdots, 2n+1)$ just defined.*

To prove this theorem we recall that the matrix (2:4) has rank $p+1$ since E_0 is normal relative to the end conditions $\psi_\mu=0$. Furthermore since E_0 satisfies the transversality condition (2:3), it follows that on E_0 the derivatives $g_{x_1}, g_{y_{i1}}, g_{x_2}, g_{y_{i2}}$ are expressible as a linear combination of the rows of the matrix (2:4), the multiplier of the first row being different from zero. The rank of the matrix (2:4) formed for the new end conditions $\psi_\rho=0$ is therefore unaltered when the elements of the first row are replaced by the derivatives $g_{x_1}, g_{y_{i1}}, g_{x_2}, g_{y_{i2}}$. The matrix thus formed is the matrix of the determinant (5:1) and has rank $2n+2$. Hence according to Theorem 2:2, E_0 is also normal relative to the end conditions $\psi_\rho=0$, and the theorem is established.

6. A fundamental sufficiency theorem. With the help of the auxiliary problem just constructed we can prove the following theorem:

THEOREM 6:1. A FUNDAMENTAL SUFFICIENCY THEOREM. *Let E_0 be an extremal arc with the following properties:*

(A) *E_0 satisfies the sufficient conditions for a proper strong relative minimum*

with respect to admissible arcs C satisfying the end conditions $\psi_\rho(C)=0$ of the auxiliary problem of Mayer defined in §5.

(B) *There is a neighborhood M of the ends of E_0 in $(x_1y_1x_2y_2)$-space such that the inequality $g(E)>g(E_0)$ holds for every extremal E of the family* (3:1) *with ends in M satisfying the conditions $\psi_\mu(E)=0$ and not identical with E_0.*

Then there exist neighborhoods $\mathfrak{F}$ of E_0 in xy-space and N of the ends of E_0 in $(x_1y_1x_2y_2)$-space such that the inequality $g(C)>g(E_0)$ holds for every admissible arc C in $\mathfrak{F}$ with ends in N satisfying the conditions $\psi_\mu(C)=0$ and not identical with E_0.

The proof is based on the following two lemmas, the proofs of which will be given in the next section.

Lemma 6:1. (Modification of Hahn's Theorem (XIV, p. 129).) *The property* (A) *for E_0 implies the existence of neighborhoods $\mathfrak{F}$ of E_0 in xy-space and M of the ends of E_0 in $(x_1y_1x_2y_2)$-space such that for every extremal E of the family* (3:1) *with ends in M the inequality $g(C)>g(E)$ holds for every admissible arc C in $\mathfrak{F}$ with ends in M satisfying the conditions $\psi_\rho(C)=\psi_\rho(E)$ and not identical with E.*

Lemma 6:2. *The property* (A) *for E_0 implies that every neighborhood M of the end values of E_0 has associated with it a second neighborhood N of these end values such that for every admissible arc C with ends in N there is an extremal E of the family* (3:1) *with ends in M satisfying the conditions $\psi_\rho(C)=\psi_\rho(E)$.*

With the help of these lemmas the proof of Theorem 6:1 is as follows. Select first neighborhoods $\mathfrak{F}$ of E_0 and M of the ends of E_0 effective as in Lemma 6:1 and as in (B). Select a second neighborhood N of the ends of E_0 related to M as in Lemma 6:2. Consider now an admissible arc C in $\mathfrak{F}$ with ends in N satisfying the conditions $\psi_\mu=0$. According to Lemma 6:2 there is an extremal E of the family (3:1) with ends in M satisfying the conditions $\psi_\mu(E)=0$, $\psi_\tau(E)=\psi_\tau(C)$, where the functions ψ_τ are those adjoined to the functions ψ_μ to form the auxiliary Mayer problem defined in §5. From Lemma 6:1 it follows that $g(C)\geqq g(E)$, and from the property (B) we have $g(E)\geqq g(E_0)$. Hence $g(C)\geqq g(E_0)$, the equality being valid only in case C coincides with E_0, as was to be proved.

7. **Proofs of two lemmas.** In order to prove Lemma 6:1 we use the result obtained by Bliss and Hestenes (XVI, p. 323)* which states that the

* In the proof of the theorem referred to here, the authors made use (XVI, Theorem 8:1) of a suggestion in an abstract by Morse, Bulletin of the American Mathematical Society, vol. 37 (1931), p. 37. Bliss and Reid proved Morse's result independently before the complete paper of Morse (XVII) appeared. Bliss and Hestenes used the proof given by Bliss, which is similar to that of Morse, and inadvertently made no reference to Morse's paper. The proof given by Morse should of course have priority.

property (A) for E_0 given in Theorem 6:1 implies the existence of a function $W(a_1, \cdots, a_n)$ such that the n-parameter family of extremals

$$(7:1)\qquad y_i = y_i(x, x_{20}, a, W_a), \quad z_i = z_i(x, x_{20}, a, W_a) \qquad (x_1 \leqq x \leqq x_2)$$

contains E_0 for parameter values (x_{10}, x_{20}, a_0) and has the determinant $|y_{ia_k}|$ different from zero along E_0. Furthermore each extremal E of the family (7:1) has on it the element $(x, y, z) = (x_{20}, a, W_a)$, where the a_i are the parameter values defining E. If now we select $n-1$ functions $W_r(a_1, \cdots, a_n)$ having continuous first and second partial derivatives and such that the determinant $|W_{a_i} W_{ra_i}|$ is different from zero for the values $a_i = a_{i0}$, then the $(2n-1)$-parameter family of extremals

$$(7:2)\qquad \begin{aligned} y_i &= y_i(x, x_{20}, a, W_a + b_r W_{ra}) = y_i(x, a, b), \\ z_i &= z_i(x, x_{20}, a, W_a + b_r W_{ra}) = z_i(x, a, b) \qquad (x_1 \leqq x \leqq x_2) \end{aligned}$$

contains E_0 for parameter values $(x_{10}, x_{20}, a_0, b=0)$. Moreover every extremal E of this family has on it the element $(x, y_i, z_i) = (x_{20}, a_i, W_{a_i} + b_r W_{ra_i})$, where the parameter values a_r, b_r are those defining E. The equations expressing this fact are the equations

$$a_i = y_i(x_{20}, a, b), \quad W_{a_i} + b_r W_{ra_i} = z_i(x_{20}, a, b),$$

and by differentiation it is found that the determinant

$$\begin{vmatrix} y_{ia_k} & y_{ib_r} & 0 \\ z_{ia_k} & z_{ib_r} & z_i \end{vmatrix}$$

is different from zero for the values $(x, a, b) = (x_{20}, a_0, 0)$. Hence the family (7:2) is one of the type (3:1), its multipliers $\lambda_\alpha(x, a, b)$ being found in the usual manner (XVI, pp. 309–311).

Since the determinant $|y_{ia_k}|$ belonging to the family (7:1) is different from zero on E_0, the determinant $|y_{ia_k}(x, a, b)|$ belonging to the family (7:2) has the same property. Hence the system of equations

$$(7:3)\qquad y_i = y_i(x, a, b)$$

has a unique solution

$$a_i = a_i(x, y, b)$$

in a neighborhood $\mathfrak{D}$ of the values (x, y, b) belonging to E_0. The functions $a_i(x, y, b)$ are continuous and possess continuous derivatives of the first two orders in the domain $\mathfrak{D}$. If now we let

$$(7:4)\qquad \begin{aligned} p_i(x, y, b) &= y_{ix}[x, a(x, y, b), b], \\ \lambda_\alpha(x, y, b) &= \lambda_\alpha[x, a(x, y, b), b], \end{aligned}$$

then according to the condition $\mathrm{II}_{\mathfrak{N}}'$ implied by the property (A) on E_0, the domain $\mathfrak{D}$ can be so restricted that at each element (x, y, b) in $\mathfrak{D}$ the inequality

$$E[x, y, p(x, y, b), \lambda(x, y, b), y'] > 0$$

holds for every admissible set $(x, y, y') \neq (x, y, p)$, where $E(x, y, p, \lambda, y')$ is the Weierstrass E-function (XVI, pp. 317, 324). Furthermore on the hyperplane $x = x_{20}$ in xy-space the Hilbert integral I^* is independent of the path when the parameters b_r are fixed (XVI, p. 323, cf. XII, p. 269). It follows that for each set b_r the region $\mathfrak{F}$ of points (x, y), whose elements (x, y, b) are all in $\mathfrak{D}$, forms a field with slope functions and multipliers defined by equations (7:4) (XVI, p. 322). We have a family of such fields depending upon the $n-1$ parameters b_r. In each field the Weierstrass E-function is >0 unless $y_i' = p_i$. Hence according to a theorem proved by Bliss and Hestenes (XVI, p. 319) there is a neighborhood M of the end values of E_0 such that every extremal E with ends in M and belonging to one of these fields furnishes a proper strong relative minimum for the function g in the class of admissible arcs C in $\mathfrak{F}$ whose ends are in M and satisfy the conditions $\psi_\rho(C) = \psi_\rho(E)$.

Lemma 6:1 will now be established completely if we show that the neighborhood M of the ends of E_0 can be restricted so that every extremal E of the family (7:2) with ends in M is a member of one of the fields just described. To do this we select a constant h so that the set $[x, y, a_i(x, y, b), b]$ with elements (x, y, b) in $\mathfrak{D}$ is the only solution of equations (7:3) satisfying the relation

$$a_i(x, y, b) - h \leqq a_i \leqq a_i(x, y, b) + h. \tag{7:5}$$

This can always be done since the solution $a_i(x, y, b)$ of equations (7:3) is isolated. We now select a constant ϵ such that the inequalities

$$|a_i - a_{i0}| < h/2,$$
$$|a_{i0} - a_i(x, y, b)| < h/2$$

hold along every extremal E of the family (7:2) with parameter values (x_1, x_2, a, b) in an ϵ-neighborhood of those belonging to E_0. The relation (7:5) now holds for every set of values (x, y, a, b) on E. It follows that $a_i = a_i(x, y, b)$, and hence E is an extremal of one of the fields just described. This completes the proof of Lemma 6:1 since according to Theorem 3:1 the neighborhood M of the ends of E_0 can be so restricted that every extremal E of the family (7:2) with ends in M has parameter values (x_1, x_2, a, b) in the ϵ-neighborhood just defined.

In order to prove Lemma 6:2 consider first the equations

$$
\begin{aligned}
&W(x_1, y_1, x_2, y_2) = 0, \\
&\psi_\rho(x_1, y_1, x_2, y_2) = m_\rho,
\end{aligned}
\tag{7:6}
$$

where W is the function defined in Theorem 3:1. As was seen in §3, the functional determinant of these equations is different from zero on E_0. Furthermore equations (7:6) are satisfied by the set $(x_1, y_1, x_2, y_2, m) = (x_{10}, y_{10}, x_{20}, y_{20}, 0)$ belonging to E_0. Hence there is a constant $h>0$ such that equations (7:6) have a unique solution

$$
\begin{aligned}
x_1 &= x_1(m), \quad y_{i1} = y_{i1}(m), \\
x_2 &= x_2(m), \quad y_{i2} = y_{i2}(m)
\end{aligned}
\tag{7:7}
$$

for all values m_ρ satisfying the relations $|m_\rho|<h$. If h is sufficiently small, then according to Theorem 3:1 every pair of points (x_1, y_1), (x_2, y_2) can be joined by an extremal of the family (3:1). Furthermore it is clear that, if necessary, the constant h can be further restricted so that every set of values (x_1, y_1, x_2, y_2) defined by equations (7:7) with $|m_\rho|<h$ is in a preassigned neighborhood M of the end values of E_0. If now we select a second neighborhood N of the end values of E_0 so that every set of values (x_1, y_1, x_2, y_2) in N satisfies the relation $|\psi_\rho(x_1, y_1, x_2, y_2)|<h$, then every admissible arc C with ends in N determines a set of values $m_\rho = \psi_\rho(C)$ satisfying the relation $|m_\rho| <h$, and these in turn determine an extremal arc E with ends in M satisfying the conditions $\psi_\rho(E) = \psi_\rho(C)$. This proves Lemma 6:2.

8. **Sufficient conditions for relative minima.** The necessary condition I is given in §2. The symbols II$_{\mathfrak{N}}$', III' will be used to denote the strengthened conditions of Weierstrass and Clebsch as defined by Bliss and Hestenes (XVI, p. 324). The symbol IV' will be used to denote the condition IV of §4 strengthened so as to exclude the equality sign. With these definitions agreed upon we can state the following theorem:

THEOREM 8:1. SUFFICIENT CONDITIONS FOR A STRONG RELATIVE MINIMUM. *Let E_0 be an admissible arc without corners and with end points determined by values x_{10}, x_{20} and satisfying the conditions $\psi_\mu=0$. If E_0 is normal relative to the end conditions $\psi_\mu=0$, is normal on every sub-interval $x_{10}x_3$ of $x_{10}x_{20}$, and satisfies the conditions* I, II$_{\mathfrak{N}}$', III', IV', *then there exist neighborhoods $\mathfrak{F}$ of E_0 in xy-space and N of the ends of E_0 in $(x_1y_1x_2y_2)$-space such that the inequality $g(C) >g(E_0)$ holds for every admissible arc C in $\mathfrak{F}$ with ends in N satisfying the conditions $\psi_\mu(C)=0$ and not identical with E_0.*

The theorem will be established if we can show that the hypotheses of the

theorem imply those of Theorem 6:1. It is easily seen from Theorem 5:1 and from the sufficiency conditions given by Bliss and Hestenes for the case $p=2n+1$ (XVI, p. 324) that E_0 is an extremal arc having the property (A) of Theorem 6:1 provided that we can show that the condition IV', as defined above, implies that the ends of E_0 are not conjugate to each other. If the ends of E_0 were conjugate then the constants dc_s in the expressions $\delta y_i = y_{ic_s} dc_s$ could be selected not all zero so that the differentials δy_i would all vanish at the ends of E_0. If we should take these constants dc_s together with the values $dx_1 = dx_2 = 0$, then the conditions $d\psi_\mu = 0$ would be satisfied and the expression (4:2) for d^2g would vanish, which would contradict the condition IV'. Hence E_0 has property (A) of Theorem 6:1.

To prove that E_0 has the property (B) of Theorem 6:1 we first note that the conditions I, III' imply the existence of a family of extremals (3:1) containing E_0 for parameter values (x_{10}, x_{20}, c_{s0}). From conditions I, IV' it follows that $dg=0$, $d^2g>0$ for every set of differentials $(dx_1, dx_2, dc_s) \neq (0, 0, 0)$ which satisfy the conditions $d\psi_\mu = 0$. But these are the conditions (XV, p. 115) which insure that $g(x_1, x_2, c_s) > g(x_{10}, x_{20}, c_{s0})$ for all sets $(x_1, x_2, c_s) \neq (x_{10}, x_{20}, c_{s0})$ satisfying the equations $\psi_\mu(x_1, x_2, c_s) = 0$ and lying in a sufficiently small ϵ-neighborhood of (x_{10}, x_{20}, c_{s0}). Furthermore since the ends of E_0 are not conjugate the matrix (3:2) has rank $2n-1$ (XVI, p. 316), and according to Theorem 3:1 there is a neighborhood M of the ends of E_0 such that every extremal with ends in M has parameter values (x_1, x_2, c_s) in the ϵ-neighborhood described above. It follows that $g(x_1, y_1, x_2, y_2) > g(x_{10}, y_{10}, x_{20}, y_{20})$ for every extremal E with ends in M satisfying the conditions $\psi_\mu(E) = 0$ and not identical with E_0. Hence E_0 has the property (B) of Theorem 6:1 and Theorem 8:1 is established.

In a similar manner sufficient conditions for a weak relative minimum for the general problem of Mayer with variable end points can be established. The argument is like that of Bliss and Hestenes (XVI, p. 325) with the help of simple modifications of Theorem 6:1 and Lemma 6:1 above. The Theorem 10:2 of Bliss and Hestenes remains valid here if we replace the phrase "preceding theorem" by "Theorem 8:1" and the equations $\psi_\rho = 0$ by $\psi_\mu = 0$. Similarly Corollary 10:1 of the paper by Bliss and Hestenes is still effective if we replace "Theorem 10:1" by "Theorem 8:1" and ψ_ρ by ψ_μ.

Bibliography

The papers listed below are a continuation of the list at the end of the paper of Bliss and Hestenes cited here as XVI.

XIII. Mayer, *Zur Aufstellung der Kriterion des Maximums und Minimums der einfachen Integrale bei variablen Grenzwerten*, Leipziger Berichte, vol. 36 (1884), pp. 99–128, vol. 48 (1896), pp. 436–465.

XIV. Hahn, *Ueber Variationsprobleme mit variablen Endpunkten*, Monatshefte für Mathematik und Physik, vol. 22 (1911), pp. 127–136.

XV. Hancock, *Theory of Maxima and Minima*, Ginn and Company, 1917.

XVI. Bliss and Hestenes, *Sufficient conditions for a problem of Mayer in the calculus of variations*, these Transactions, vol. 35 (1933), pp. 305–326.

XVII. Morse, *Sufficient conditions in the problem of Lagrange with fixed end points*, Annals of Mathematics, (2), vol. 32 (1931), pp. 567–577.

University of Chicago,
Chicago, Ill.

THE STRUCTURE OF THE NUMBER OF REPRESENTATIONS FUNCTION IN A BINARY QUADRATIC FORM*

BY

GORDON PALL

This paper contains, primarily, the extension to any integral, binary quadratic form of the results of a recent article† concerning positive, binary, quadratic forms. With suitable conventions almost all the results carry over without change, though some of the proofs need slight alterations. Incidentally, there are treated automorphs of binary quadratic forms, and (rather fully) properties of sets of representations (representations equivalent through automorphic transformations) in a binary quadratic form.

1. Dirichlet‡ has already in all essentials extended the notion of number of representations to indefinite forms. We shall utilize the following equivalent definition.

Two representations (x,y) and (x',y') of m in the form $f=[a,b,c]$,§ that is, two integral solutions of

$$ax^2+bxy+cy^2=m, \tag{1}$$

will be called *equivalent* if they are transformable one into the other by integral automorphs of f. The class of all representations equivalent to a given one will be called a *set* of representations. The number of sets of representations of m in f will be denoted by $f(m)$. (In MZ, $f(m)$ denoted the number of representations of m in f.)

This definition becomes more interesting when we observe that, if $d(=b^2-4ac)>0$, the number of sets of representations of m in $[a,b,c]$ is equal to the actual number of solutions of (1) together with certain inequalities (cf. Theorem 9). The writer developed the theory of these inequalities before noticing that Dirichlet (§87, loc. cit.) obtains one such system. However, we shall obtain a substantial improvement on Dirichlet's inequalities and give a more complete discussion of the infinitely many alternative systems. The treatment in §3 is fairly comprehensive.

* Presented to the Society, October 29, 1932; received by the editors September 26, 1932.

† Mathematische Zeitschrift, vol. 36 (1933); this article will be referred to here as MZ.

‡ Cf. §§86 and 87 of *Vorlesungen über Zahlentheorie*, 4th edition, 1894.

§ We use Kronecker forms, for simplicity, throughout. Hence $[a,b,c]$ stands for $ax^2+bxy+cy^2$. For the automorphs see §2.

It will then be observed that, once the least positive solution t_1, u_1 of

$$t^2 - du^2 = 4 \tag{2}$$

is known, the labor of representing a number *by trial* in $ax^2+bxy+cy^2$ is, when $d>0$, on a par with the work when $d<0$. For example, for either of the equations

$$x^2 + 2y^2 = n, \quad x^2 - 2y^2 = n,$$

where n is a given positive integer, we need to try only the values y^2 such that $0 \leqq y^2 \leqq n/2$ to obtain a representation in every set. (For $x^2-2y^2=n$ the inequalities of Dirichlet require us to examine $0 \leqq y^2 < 4n$.)

To obtain this improvement it is necessary to introduce a convention whereby solutions on the boundaries of the inequalities count as $\frac{1}{2}$ instead of 1 (cf. Theorem 9). For example, if $d=8$, and $f=[1, 0, -2]$, then $f(n)$ is equal to the number $g(n)$ of integral solutions (x, y) of

$$n = x^2 - 2y^2,\ 0 \leqq y \leqq (n/2)^{1/2} (n > 0),\ (-n/2)^{1/2} \leqq y \leqq (-n)^{1/2} (n < 0), \tag{3}$$

except when $\pm n = k^2$ or $2k^2$ (k integral) in which cases $f(n)=g(n)-1$. The condition of inequality may be replaced by $|x| \geqq 2y \geqq 0$, or $|y| \geqq x \geqq 0$.

If $d=b^2-4ac$ is negative or is a positive square, the number w of integral automorphs of f is 2, except that w is 4 if $d=-4$ and w is 6 if $d=-3$. Then the number of representations of n in f is $wf(n)$.

Unless otherwise specified each (binary) *form* in the sequel is a *primitive, integral, binary, quadratic form of discriminant d, where d is a non-zero integer* $\equiv 0$ *or* 1 (mod 4). For simplicity we do *not* make the usual convention that the forms are positive if $d<0$.

For any d there are a finite number h of (primitive) classes of forms, say $C_0, C_1, \cdots, C_{h-1}$. Representative forms from these classes are denoted, respectively, by $f_0, f_1, \cdots, f_{h-1}$. We shall always take C_0 to be the principal class, which represents $+1$. The system of representative forms will be designated by $S(=S_d)$. The sum of the numbers of sets of representations of n in the h forms will be denoted by $S(n)$, so that

$$S(n) = f_0(n) + f_1(n) + \cdots + f_{h-1}(n). \tag{4}$$

The system of classes C_i constitutes under composition a finite abelian group with C_0 as identity element. We shall assume their behavior in this respect as known, and shall interpret C_iC_j, etc., as the product classes under composition. Further if f is a form belonging to a class F, f^{-1} will denote the opposite form (belonging to the reciprocal class F^{-1}); and if g belongs to G, then f^rg^s will denote any form of the class F^rG^s.

An ambiguous class C is characterized by the equation $C^2=C_0$; or by $C=C^{-1}$; or by containing a form $[a, b, c]$ in which $a|b$.

We can now state our principal results.

The function $S(n)$ is a factorable function; for any relative-prime integers n_1 and n_2,

$$S(n_1n_2) = S(n_1)S(n_2). \tag{5}$$

An integer n is *semiprime to d*, by definition, if n is divisible by no prime p such that

$$p>2 \text{ and } p^2\,|d, \text{ or } p=2 \text{ and } d\equiv 0 \text{ or } 4 \pmod{16}. \tag{6}$$

For any n semiprime to d, we shall prove

$$S(n) = \sum_{\nu|n}(d\,|\,\nu), \tag{7}$$

where ν ranges over the positive divisors of n, and $(d|\nu)$ is the Kronecker symbol.

For example in (3), the number $g(n)$ of integral solutions (x, y) is given by $g(n)=\sum(2|\nu)$ unless $\pm n$ is a square or the double of a square, and then $g(n)=1+\sum(2|\nu)$.

The system $\{f_0(n), \cdots, f_{h-1}(n)\}$ is *reducible* in the following sense: for every prime p not satisfying (6) and every integer $a>0$ there exists a matrix of h^2 numbers $\psi_{ij}(p, a)$ $(i, j=0, \cdots, h-1)$ such that, for every integer m prime to p,

$$f_i(p^am) = \sum_{j=0}^{h-1}\psi_{ij}(p, a)f_j(m). \tag{8}$$

More precisely the following formulas hold.

Let f be a form of discriminant d and F its class.

If $p|d$ but does not satisfy (6), p is represented by an ambiguous class C of discriminant d. If g belongs to C^aF,

$$f(p^an) = g(n) \text{ for every integer } n. \tag{9}$$

If $(d|p)=-1$ (Kronecker symbol),

$$f(pm) = 0, \quad f(p^2n) = f(n), \tag{10}$$

for every m prime to p and every integer n.

If $(d|p)=1$ there is a form g of discriminant d representing p. For every integer n,

$$f(pn) + f(n/p) = fg(n) + fg^{-1}(n). \tag{11}$$

Solving this relation as in MZ §6 we obtain

$$f(p^a m) = \sum_{\alpha=0}^{a} fg^{a-2\alpha}(m), \tag{12}$$

which holds for every integer $a \geqq 0$ and m prime to p.

Finally let p satisfy (6). Again $f(pm) = 0$ if p does not divide m. Now there exists a form g, of discriminant $d' = d/p^2$, which may be characterized as representing every number represented by f. For this form,

$$f(p^2 n) = \sigma g(n) \qquad \text{(every } n\text{)}, \tag{13}$$

where σ is partially defined by

$$\begin{cases} \sigma = 1 \text{ if } d' < -4 \text{ or } d' \text{ is a square,} \\ \sigma = 2 \text{ if } d' = -4, \\ \sigma = 3 \text{ if } d' = -3.* \end{cases} \tag{14}$$

If d' is positive but not square, employ the notation t_k, u_k for the successive solutions of

$$t^2 - d'u^2 = 4, \tag{15}$$

t_1, u_1 being the least positive solution (as in §3.4 with d' in place of d). Then σ is the (least) positive-index such that

$$u_\sigma \equiv 0, \quad u_k \not\equiv 0 \quad (0 < k < \sigma) \qquad (\text{mod } p). \tag{16}$$

To understand these formulas properly we should observe that

(17) If p is a prime, p and $-p$ are each represented in one of the classes of S unless (6) holds or $(d|p) = -1$;

(18) Either is represented in at most one class and the reciprocal class.

A class and its reciprocal, being improperly equivalent, represent the same numbers. If $d < 0$ the classes of S will occur in pairs, each class being accompanied by its negative.

The generality of our results as holding even for d square (but $\neq 0$) may be emphasized.

All the results of MZ, with the slight changes obvious from the preceding statements, hold for any integral binary quadratic form with discriminant $d \neq 0$. If $d < 0$ the form $-f_0$ of exponent 2 may be adjoined to the basis of MZ §1, if desired.

It is interesting to obtain a formula for the ψ_{ij} of (8). We can choose

---

* The reader will note that if $f(n)$ meant the actual number of representations instead of the number of sets we should have $f(p^2 n) = g(n)$ in all the cases in (14).

$\psi_{ik}=\psi_{ij}$ if f_j and f_k belong to reciprocal classes. Then $\psi_{ij}=\frac{1}{2}f_i(p^a q)$, where q is any prime represented in f_j (and f_k). We may choose q so that $q\neq p$ and $(d|q)=1$. Hence

$$\psi_{ij}(p, a) = \tfrac{1}{2}\{f_i f_j(p^a) + f_i f_k(p^a)\} \quad (\text{where } f_j f_k = f_0). \tag{19}$$

Suppose $(d|p)=1$, $g(p)>0$. Then $2\psi_{ij}(p, a)$ is the number of the elements of the sequence $g^a, g^{a-2}, \cdots, g^{-a}$ belonging to the class of $f_i f_j$ or its reciprocal, plus the number in the class of $f_i f_k$ or its reciprocal.

2.1. **Automorphs.** We prove the following theorem.

THEOREM 1. *Let a, b, c be integers of g.c.d. unity, set $d=b^2-4ac$ and suppose $d\neq 0$. Then all integral automorphs (of determinant +1) of $[a, b, c]$ are given by*

$$\begin{aligned} x &= \tfrac{1}{2}(t - bu)x_0 - cuy_0, \\ y &= aux_0 + \tfrac{1}{2}(t + bu)y_0, \end{aligned} \tag{20}$$

as (t, u) ranges over all integral solutions of

$$t^2 - du^2 = 4. \tag{21}$$

Let I denote the identity matrix. If T is a non-singular matrix, $T^{-1}IT=I$. Hence we have the following lemma:

LEMMA 1. *If a form has only the two automorphs with matrices $\pm I$, the same is true of all equivalent forms.*

First let d be a positive square. Set $d=\Delta^2$, $\Delta>0$. Then (21) has only the trivial solutions $(\pm 2, 0)$ (for which (20) has matrices $\pm I$). The theorem will therefore follow if it holds for a form equivalent to $f=[a, b, c]$. Now f is equivalent to one and only one of the $\phi(\Delta)$ forms

$$[k, \Delta, 0],\ 0 \leqq k < \Delta,\ k \text{ prime to } \Delta, \tag{22}$$

and it is easy to show that these have only the two trivial automorphs.*

* For a primitive form $(\lambda x+\mu y)(\nu x+\rho y)$ is equivalent to a form $x(\sigma x+\tau y)$, where, by the discriminant, $\tau=\pm\Delta$. Replacing y by $y+\kappa x$ we can alter σ by multiples of Δ. Evidently $[k, \Delta, 0]\sim'[k, -\Delta, 0]$, where $\sim'$ means "is improperly equivalent to". Hence the fact that (22) constitutes a complete representative system of forms will follow once we prove that $[k, \Delta, 0]\sim'[l, \Delta, 0]$ when $kl\equiv 1 \pmod{\Delta}$, and that $[k, \Delta, 0]\sim[l, \Delta, 0]$ only when $k\equiv l \pmod{\Delta}$. To prove both these facts and to obtain all transformations carrying $[k, \Delta, 0]$ into $[l, \Delta, 0]$, compare coefficients in the identity

$$(\alpha x+\beta y)[k(\alpha x+\beta y)+\Delta(\gamma x+\delta y)]=x(lx+\Delta y),$$

where k and l are given prime to Δ, $0\leqq k<\Delta$, $0\leqq l<\Delta$. Then either $\beta=0$ or $k\beta+\Delta\delta=0$. The former case leads to $\alpha\delta=1$, $k\pm\Delta\gamma=l$, whence $k=l$ and $\gamma=0$ and the transformation matrix is $\pm I$. The latter case leads to

$$\alpha\delta-\beta\gamma=-\beta(\alpha k+\Delta\gamma)/\Delta=-1,\quad \beta=\pm\Delta,\quad \alpha=\pm l,\quad \delta=\mp k,\quad \gamma=\mp(kl-1)/\Delta.$$

This footnote and formula (22) will be useful to the reader who may wish to verify that properties which he knows to hold for d not square continue to hold when d is a square$\neq 0$.

Second we shall indicate a uniform proof, valid at least when d is not a square, by a modification of Dickson's *Introduction to the Theory of Numbers*, §§60 and 69. By taking $R^2=d$, $R>0$ if $d>0$, $-iR>0$ if $d<0$, we can define first and second roots in §60 for any form with $da\neq 0$. Then Theorems 72 and 73 hold unchanged together with their proofs, at least if d is not square. Also §69 holds for any non-square d.

2.2. **Proper sets of representations.** If (x, y) and (x_0, y_0) are related by (20) we say that (x, y) and (x_0, y_0) are equivalent representations in f. As already defined, all (x, y) equivalent to a given one comprise a *set*.

THEOREM 2. *The g.c.d. of x and y is the same for all (x, y) of a set.*

This is evident on solving (20) for x_0 and y_0. We may now speak of *proper sets*, that is, sets in which the g.c.d. is 1.

2.3. **Proper sets and the congruence $z^2\equiv d$ (mod $4m$).** Let Σ denote the aggregate of solutions z of

$$z^2 \equiv d \pmod{4m},\ 0 \leqq z < 2\,|\,m\,|\,, \tag{23}$$

such that

$$z,\ m, \text{ and } (z^2 - d)/(4m) \text{ have g.c.d. } 1. \tag{24}$$

We shall set up a (1, 1) correspondence between the elements z of Σ and the various proper sets of representations of m by the h forms in S.

For any z write $z^2-d=4ml$. Then $\phi=[m, z, l]$ is a primitive form of discriminant d, and hence is equivalent to just one of the forms of S, say to $f=[a, b, c]$. For brevity we shall write

$$T = \begin{pmatrix} x & \xi \\ y & \eta \end{pmatrix},\quad A = \begin{pmatrix} \frac{1}{2}(t - bu) & -cu \\ au & \frac{1}{2}(t + bu) \end{pmatrix}. \tag{25}$$

If T is the matrix of one transformation of determinant $+1$ carrying f into ϕ, then the totality of such matrices is given by AT as t, u range through all integral solutions of (21). But then x and y are relative-prime and $m=ax^2+bxy+cy^2$, that is, (x, y) is a proper representation of m in f. And the class of first columns of the matrices AT is a set of representations of m in f.

Conversely, let (x, y) be a proper representation of m in f. We can choose integers η and ξ so that

$$x\eta - y\xi = 1, \tag{26}$$

and then the general form of such integers is $\xi+tx$, $\eta+ty$, where t is an integer. On applying the transformation with matrix T to f we derive $\phi=[m, n, l]$ where $m=ax^2+bxy+cy^2$ and

$$n = 2ax\xi + b(x\eta + y\xi) + 2cy\eta, \tag{27}$$

while l is determined by the discriminant. If we replace ξ and η in T by $\xi+tx$ and $\eta+ty$, m is unchanged but n is replaced by $n+2tm$. If we replace (x, y) by an equivalent representation, so that T is replaced by AT or a parallel matrix,* we again derive ϕ or a parallel form. Thus the (1, 1) correspondence is established.

THEOREM 3. *Let a, b, c have g.c.d. 1, let $d=b^2-4ac\neq 0$, and let m be an integer $\neq 0$. Let $f'(m)$ denote the number of proper sets of representations of m in $f=[a, b, c]$. Then $f'(m)$ is equal to the number of roots z of (23) such that $[m, z, (z^2-d)/(4m)]$ is equivalent to f.*

2.4. **On $S'(n)$ and $S(n)$.** Evidently $f(m)=\sum f'(m/q^2)$, where q^2 ranges over the square divisors of m. The number of proper sets of representations of m in S is

$$S'(m) = f_0'(m) + \cdots + f_{h-1}'(m), \tag{28}$$

and is equal to the number of solutions z of (23) and (24). Also,

$$S(m) = \sum S'(m/q^2). \tag{29}$$

Proceeding as in MZ but now allowing n to be negative as well as positive, we see that $S'(n)$ and $S(n)$ are factorable (MZ, §2).† Further we have

$$S'(1) = S'(-1) = S(1) = S(-1) = 1, \tag{30}$$

$$S'(n) = S'(-n), \quad S(n) = S(-n) \quad (\text{every } n). \tag{31}$$

For any prime $p\geqq 2$ we have, using the Kronecker symbol $(d\,|\,p)$,

$$S'(p^a) = 1 + (d\,|\,p) \qquad \text{if } p \text{ does not divide } d, \tag{32}$$

$$\begin{aligned} S(p^a) &= a + 1 && \text{if } (d\,|\,p) = 1,\\ &= \tfrac{1}{2}\{1 + (-1)^a\} && \text{if } (d\,|\,p) = -1,\\ &= 1 && \text{if } p\,|\,d \text{ but (6) does not hold.}\end{aligned} \tag{33}$$

If p does not satisfy (6) we have therefore

$$S(p^a) = 1 + (d\,|\,p) + \cdots + (d\,|\,p^a). \tag{34}$$

Hence, if n is semiprime to d,

$$S(n) = \sum (d\,|\,\nu) \tag{35}$$

summed for the positive divisors ν of n. In calculating $S(n)$ it is generally simpler to factor n into primary components, $n=\pm\Pi p^a$, and to employ (33).

* Two matrices like T are called parallel if their first columns are identical and their second columns differ by an integral multiple of their first columns.

† $S(n)$ and $S'(n)$ are used here instead of $r(n)$ and $r'(n)$ in MZ. The value of $S(p^a)$ for all cases may be read from the table of values of $r(p^a)$ in §3 of MZ.

2.5. **Representation of p or $-p$.** Let $m=\pm p$ where p is a prime. The number of roots of (23) and (24) is 0 if $(d|p)=-1$ or if (6) holds, 1 if $p|d$ but (6) does not hold, and 2 if $(d|p)=1$. In the second case the root of (23) is $z=0$ if d is even and p is odd, or if $d\equiv 8 \pmod{16}$ and $p=2$; the root is $z=p$ if d is odd, or if $d\equiv 12 \pmod{16}$ and $p=2$; whence the form

$$[\pm p, z, \cdots]$$

associated is ambiguous. In the third case the two roots are of the forms z, $2p-z$, where $0<z<p$, and the classes which represent p are represented by the two forms

$$[\pm p, z, \cdots], [\pm p, 2p-z, \cdots],$$

and are improperly equivalent. These facts prove (17), (18), and the following theorem.

THEOREM 4. *Let $m=\pm p$, p prime. Let f denote a primitive form representing m, and let F be the class of f. Then*

$$f(m)=2 \text{ if } p \text{ does not divide } d \text{ and } F \text{ is ambiguous}; \tag{36}$$
$$=1 \text{ if } p|d \text{ or if } p \text{ does not divide } d \text{ and } F \text{ is not ambiguous}.$$

In case $p|d$, F is necessarily ambiguous.

2.6. **Representation of $\pm p_1p_2$, $(d|p_i)=1$.** As in MZ §4 we may prove the following result.

THEOREM 5. *Let p_1 and p_2 be distinct primes such that $(d|p_1)=(d|p_2)=1$. Let ϵ_1 and ϵ_2 be signs $+$ or $-$. Let g_i represent ϵ_ip_i $(i=1, 2)$. Let G_i be the class of g_i $(i=1, 2)$. Let f denote any form of the product class G_1G_2. Then f represents $m=\epsilon_1\epsilon_2p_1p_2$, and*

(37) *$f(m)=4$, 2 or 1 according as G_1G_2 coincides with all, just one, or none of the classes $G_1G_2^{-1}$, $G_1^{-1}G_2$, $G_1^{-1}G_2^{-1}$.**

3. **Some properties of sets of representations.** By definition, (x, y) and (x_0, y_0) are equivalent representations in a form $[a, b, c]$, or belong to the same set of representations in $[a, b, c]$ if solutions of (2) exist satisfying (1). If there is no ambiguity as to the form involved we may write $(x, y)\sim(x_0, y_0)$.

3.1. **Transformation of sets.** We prove the following theorem.

THEOREM 6. *Let $f=[a, b, c]$ and $g=[a', b', c']$ be primitive integral forms of discriminants d and $d\epsilon^2$ respectively, $d\epsilon\neq 0$. If*

$$x=\alpha x'+\beta y',\quad y=\gamma x'+\delta y',\quad \alpha, \beta, \gamma, \delta \text{ integers}, \tag{38}$$

* I.e., according as all, just one of, or none of G_1, G_2, and G_1G_2 are ambiguous.

where $\alpha\delta-\beta\gamma=\epsilon$, *is a transformation of* f *into* g, *and if* (x', y') *and* (x_0', y_0') *are equivalent representations in* g, *then*

$$(x, y) = (\alpha x' + \beta y', \gamma x' + \delta y'), \ (x_0, y_0) = (\alpha x_0' + \beta y_0', \gamma x_0' + \delta y_0') \tag{39}$$

are equivalent representations in f.

For we have

$$\begin{aligned} a' &= a\alpha^2 + b\alpha\gamma + c\gamma^2, \quad b' = 2a\alpha\beta + b(\alpha\delta + \beta\gamma) + 2c\gamma\delta, \\ c' &= a\beta^2 + b\beta\delta + c\delta^2. \end{aligned} \tag{40}$$

By assumption, integers t', u' exist satisfying

$$t'^2 - d\epsilon^2 u'^2 = 4 \tag{41}$$

and such that

$$\begin{aligned} x' &= \tfrac{1}{2}(t' - b'u')x_0' - c'u'y_0', \\ y' &= a'u'x_0' + \tfrac{1}{2}(t' + b'u')y_0'. \end{aligned} \tag{42}$$

Using (39) and (40) it is easy to verify that (20) holds with

$$t = t', \ u = u'\epsilon. \tag{43}$$

3.2. **Opposite and ambiguous sets.** It is plain from (20) that either of the relations

$$a \mid by, \ c \mid bx \tag{44}$$

holds for all or none of the elements (x, y) of a set. We call such a relation an invariant of the set. Another example was the g.c.d. of §2.2.

If $a \mid by$ then with each integral solution (x, y) of

$$ax^2 + bxy + cy^2 = n \tag{45}$$

is associated a solution (x', y) where $x' = -by/a - x$. It is easy to verify that if (x, y) and (x_0, y_0) are related by (20), then

$$x' = \tfrac{1}{2}(t + bu)x_0' + cuy_0, \ y = -aux_0' + \tfrac{1}{2}(t - bu)y_0, \tag{46}$$

where $x_0' = -by_0/a - x_0$. Hence (x', y) and (x_0', y_0) are in the same set of solutions in $[a, b, c]$. The set thus associated with a given one in which $a \mid by$ will be called the *x-opposite set.* A set in which $a \mid by$, and which coincides with its *x-opposite set,* will be called *x-ambiguous.* We shall see later that a set is x-ambiguous if and only if it contains an element (x, y) in which dy^2 has one of the values

$$0, \ -4an, \ (t_1 - 2)an, \ -(t_1 + 2)an. \tag{47}$$

Similarly, if $c\mid bx$ in a set, there is associated a *y-opposite* set of representations (x, y') in $[a, b, c]$, $y' = -bx/c - y$. A *y-ambiguous* set is one which coincides with its y-opposite set. Later we shall prove that a set is y-ambiguous if and only if it contains an element (x, y) in which dx^2 has the value

$$0, \ -4cn, \ (t_1 - 2)cn, \quad \text{or} \ -(t_1 + 2)cn. \tag{48}$$

Excluding from consideration the trivial set which consists of the single element $(0, 0)$ we have the following theorem.

THEOREM 7. *Let a, b, c be relative-prime integers, $ac\neq 0$, and let both $a\mid by$ and $c\mid bx$ hold for the elements of a set. Then the two opposite sets coincide if and only if $ac\mid b$.*

For in order that they should coincide it is necessary and sufficient that for each (or some) element (x, y) of the set there shall exist integers t, u satisfying (21) and

$$\begin{aligned} -by/a - x &= \tfrac{1}{2}(t - bu)x - cu(-bx/c - y), \\ y &= aux + \tfrac{1}{2}(t + bu)(-bx/c - y). \end{aligned} \tag{49}$$

Multiply (49_1) by $\frac{1}{2}(t+bu)$, (49_2) by cu, and add, obtaining the first of the following equations, the second being a rearrangement of (49_2):

$$\begin{aligned} ax\{1 + \tfrac{1}{2}(t + bu)\} &= y\{acu - \tfrac{1}{2}b(t + bu)\}, \\ cy\{1 + \tfrac{1}{2}(t + bu)\} &= x\{acu - \tfrac{1}{2}b(t + bu)\}. \end{aligned} \tag{50}$$

Now if $ac\mid b$, the integers $t = (b^2 - 2ac)/(ac)$, $u = -b/(ac)$ satisfy (21) and (49). Hence it remains only to show that (50) implies that $ac\mid b$.

If, in (50), $t + bu = -2$, we have

$$y(acu + b) = 0 = x(acu + b),$$

whence $u = -b/(ac)$ is an integer. Suppose $t + bu \neq -2$. Then (50) implies

$$ax^2 = cy^2. \tag{51}$$

If now ac does not divide b, (51) will have to be satisfied by every element (x, y) of the set, which is impossible unless the number of elements in the set is 2 (or the set contains both (x, y) and $(x, -y)$, whence $b = 0$). Finally let the set consist of two elements (x, y) and $(-x, -y)$. If the two opposite sets coincide, either $-x - by/a = x$ and $y = -y - bx/c$, or $-x - by/a = -x$ and $y = y + bx/c$; the first case is impossible and the second implies $b = 0$.

If $a\mid by$ in a set, it is x-ambiguous if and only if for each (or for some) (x, y) of the set there exists a solution (t, u) of (21) such that

$$\begin{aligned} 2ax + by &= \tfrac{1}{2}t(-2ax - by) + \tfrac{1}{2}duy, \\ y &= \tfrac{1}{2}u(-2ax - by) + \tfrac{1}{2}ty. \end{aligned} \tag{52}$$

These may be combined into the single equation (cf. (21))

$$\frac{2ax+by}{y}=\frac{t-2}{u}\left(=\frac{du}{t+2}\right), \tag{53}$$

where $u=0$, $t=2$ is equivalent to $2ax+by=0$, and $u=0$, $t=-2$ is interpreted as $y=0$.

Similarly a set is y-ambiguous if and only if $c\mid bx$ and for each (or some) (x, y) of the set there exists a solution (t, u) of (21) and

$$\frac{2cy+bx}{x}=\frac{t-2}{-u}\left(=\frac{-du}{t+2}\right). \tag{54}$$

Here $(t, u)=(\pm 2, 0)$ corresponds to $2cy+bx=0$ or $x=0$.

3.3. **A congruencial property of sets mod p.** Let a, b, c be relative prime integers, $d=b^2-4ac\neq 0$. Let p be any prime not dividing ad and such that $(d\mid p)=1$ (Kronecker symbol). Then there are just two distinct roots m_1, m_2 of

$$am^2+bm+c\equiv 0 \pmod p,\quad 0\leqq m<p. \tag{55}$$

For any integer n, each integral solution (x, y) of

$$ax^2+bxy+cy^2=pn \tag{56}$$

satisfies one and only one of the three conditions

$$\begin{aligned} &x\equiv m_1y,\ y\not\equiv 0;\quad x\equiv m_2y,\ y\not\equiv 0;\\ &x\equiv y\equiv 0 \qquad\qquad \pmod p. \end{aligned} \tag{57}$$

It is easy to verify that each of the conditions (57_1), (57_2), (57_3) is an invariant property of any set of solutions (x, y) of (56).

As regards (57_3) this fact is evident from Theorem 2. Assume in (20) that $x_0\equiv m_iy_0 \pmod p$. We readily deduce $x\equiv m_iy \pmod p$.

3.4. **Concerning the equation $t^2-du^2=4$.** For the remainder of §3 we shall assume that *d is positive but not square*. Accordingly $d\geqq 5$.

All solutions (t, u) of (21) in integers are (t_k, u_k) and $(-t_k, -u_k)$, where $t_0=2$ and $u_0=0$, (t_1, u_1) is the "least positive" solution of (21), while the remaining solutions are linked together by the equations

$$\begin{aligned} 2t_{k+l}&=t_lt_k+du_lu_k,\\ 2u_{k+l}&=u_lt_k+t_lu_k, \end{aligned} \tag{58}$$

valid for all integers k and l.

Hence it is readily proved that

$$
\begin{aligned}
&t_{-k} = t_k, \quad u_{-k} = -u_k,\\
(59)\qquad &2 < t_1 < t_2 < \cdots, \quad 0 < u_1 < u_2 < \cdots,\\
&\frac{t_k}{u_k} > \frac{t_k + t_{k+1}}{u_k + u_{k+1}} > \frac{t_{k+1}}{u_{k+1}}, \quad \frac{t_k}{u_k} \to d^{1/2} \qquad (k = 0, 1, 2, \cdots).
\end{aligned}
$$

Here and later t_0/u_0 is interpreted as $+\infty$.

It will be useful to note the relations

$$
(60)\qquad \frac{t_h + t_i}{u_h + u_i} = \frac{t_{h+j} + t_{i-j}}{u_{h+j} + u_{i-j}}, \quad \frac{t_h - t_i}{u_h - u_i} = \frac{t_{h+j} - t_{i-j}}{u_{h+j} - u_{i-j}}
$$

which hold for all integers h, i, j such that the denominators are different from zero. To prove these formulas, cross-multiply and use (58_2) and (59_1). The second result is related to the first since

$$
(61)\qquad \frac{t_h + t_i}{u_h + u_i} \cdot \frac{t_h - t_i}{u_h - u_i} = d.
$$

Since $t_0 = 2$ and $u_0 = 0$, we have, with obvious conventions for the cases where the denominators vanish,

$$
(62)\qquad
\begin{aligned}
&\frac{t_{2k} + 2}{u_{2k}} = \frac{t_k}{u_k}, \quad \frac{t_{2k+1} + 2}{u_{2k+1}} = \frac{t_{k+1} + t_k}{u_{k+1} + u_k},\\
&\frac{t_{2k} - 2}{u_{2k}} = \frac{du_k}{t_k}, \quad \frac{t_{2k+1} - 2}{u_{2k+1}} = \frac{d(u_{k+1} + u_k)}{t_{k+1} + t_k}.
\end{aligned}
$$

3.5. Distribution of the solutions in a set relative to t_k, u_k. Let (x_0, y_0) denote any given solution in a set of solutions of (45). Rearranging (20) slightly we see that the aggregate of solutions (x, y) of the set are given by (x_k, y_k) and $(-x_k, -y_k)$ $(k = 0, \pm 1, \pm 2, \cdots)$, where

$$
(63)\qquad
\begin{aligned}
2ax_k + by_k &= \tfrac{1}{2}(2ax_0 + by_0)t_k + \tfrac{1}{2}dy_0u_k,\\
y_k &= \tfrac{1}{2}(2ax_0 + by_0)u_k + \tfrac{1}{2}y_0t_k.
\end{aligned}
$$

A pair of equations equivalent to (63) is

$$
(64)\qquad
\begin{aligned}
2cy_k + bx_k &= \tfrac{1}{2}(2cy_0 + bx_0)t_{-k} + \tfrac{1}{2}dx_0u_{-k},\\
x_k &= \tfrac{1}{2}(2cy_0 + bx_0)u_{-k} + \tfrac{1}{2}x_0t_{-k}.
\end{aligned}
$$

It is convenient to write

$$
(65)\qquad X_k = 2ax_k + by_k, \quad Y_k = 2cy_k + bx_k.
$$

Hence we have

$$(66)\qquad \begin{aligned} 2y_{-1} &= y_0t_1 - X_0u_1, & 2X_{-1} &= X_0t_1 - dy_0u_1,\\ 2x_{-1} &= x_0t_1 + Y_0u_1, & 2Y_{-1} &= Y_0t_1 + dx_0u_1. \end{aligned}$$

Substituting for t_k and u_k from $t_{k+1}=\frac{1}{2}t_1t_k+\frac{1}{2}du_1u_k$ and $u_{k+1}=\frac{1}{2}u_1t_k+\frac{1}{2}t_1u_k$, and employing (66) we obtain the following four systems each equivalent to (63) or (64):

$$(67)\qquad u_1X_k = -y_{-1}t_k + y_0t_{k+1},\qquad u_1y_k = -y_{-1}u_k + y_0u_{k+1};$$

$$(68)\qquad u_1X_k = -X_{-1}u_k + X_0u_{k+1},\qquad du_1y_k = -X_{-1}t_k + X_0t_{k+1};$$

$$(69)\qquad u_1Y_k = x_{-1}t_k - x_0t_{k+1},\qquad u_1x_k = -x_{-1}u_k + x_0u_{k+1};$$

$$(70)\qquad u_1Y_k = -Y_{-1}u_k + Y_0u_{k+1},\qquad du_1x_k = Y_{-1}t_k - Y_0t_{k+1}.$$

Since $X_k^2-dy_k^2=4an$, we have $|X_k|>R|y_k|$ if $an>0$, and $R|y_k|>|X_k|$ if $an<0$, where $R=d^{1/2}$. Since also $t_k>Ru_k$ it is evident from (63) that X_k has the same sign as X_0 if $an>0$, and that y_k has the same sign as y_0 if $an<0$. Similarly from (64), Y_k has the same sign as Y_0 if $cn>0$, and x_k has the same sign as x_0 if $cn<0$.

From (67)-(70) with $k=0$ or -1 we have, on performing certain subtractions,

$$\begin{aligned} du_1(y_0 - y_{-1}) &= (t_1 - 2)(X_{-1} + X_0),\\ u_1(X_0 - X_{-1}) &= (t_1 - 2)(y_{-1} + y_0),\\ -u_1(Y_0 - Y_{-1}) &= (t_1 - 2)(x_{-1} + x_0),\\ -du_1(x_0 - x_{-1}) &= (t_1 - 2)(Y_{-1} + Y_0). \end{aligned}$$

But (x_0, y_0) may be any element of the set. Hence, incorporating the preceding result, we have the following:

(71)
if $an>0$, every X_k and y_k-y_{k-1} has the same sign;
if $an<0$, every y_k and X_k-X_{k-1} has the same sign;
if $cn>0$, every Y_k and $x_{k-1}-x_k$ has the same sign;
if $cn<0$, every x_k and $Y_{k-1}-Y_k$ has the same sign.*

Hence we can choose an element (x_0, y_0) of the set such that (for every integer k)

(72) $X_k>0$ and $y_{-1}<0 \leqq y_0$ if $an>0$; $y_k>0$ and $X_{-1}<0 \leqq X_0$ if $an<0$;

or such that

(73) $Y_k<0$ and $x_{-1}<0 \leqq x_0$ if $cn>0$; $x_k<0$ and $Y_{-1}<0 \leqq Y_0$ if $cn<0$.

---

* It is easy to deduce from (71), if $a>0$, $b\geqq 0$, $c<0$, that
if $n>0$, every X_k, x_k, y_k-y_{k-1}, $Y_{k-1}-Y_k$ has the same sign as x_0;
if $n<0$, every Y_k, y_k, $x_{k-1}-x_k$, X_k-X_{k-1} has the same sign as y_0.

For this choice of (x_0, y_0) in case (72) we write

(74) $$\lambda = -y_0/y_{-1} \text{ if } an > 0,\quad \lambda = -X_0/X_{-1} \text{ if } an < 0;$$

and in case (73) we write

(75) $$\lambda = -x_0/x_{-1} \text{ if } cn > 0,\quad \lambda = -Y_0/Y_{-1} \text{ if } cn < 0.$$

Thus $\lambda \geqq 0$.

In these respective four cases, we have by (59_1) and (67)-(70) the following results:

(76) $$\begin{aligned} \frac{X_k}{y_k} &= \frac{t_k + \lambda t_{k+1}}{u_k + \lambda u_{k+1}}, & \frac{X_{-1-k}}{-y_{-1-k}} &= \frac{t_{k+1} + \lambda t_k}{u_{k+1} + \lambda u_k}; \\ \frac{X_k}{y_k} &= \frac{d(u_k + \lambda u_{k+1})}{t_k + \lambda t_{k+1}}, & \frac{-X_{-1-k}}{y_{-1-k}} &= \frac{d(u_{k+1} + \lambda u_k)}{t_{k+1} + \lambda t_k}; \end{aligned}$$

(77) $$\begin{aligned} \frac{-Y_k}{x_k} &= \frac{t_k + \lambda t_{k+1}}{u_k + \lambda u_{k+1}}, & \frac{-Y_{-1-k}}{-x_{-1-k}} &= \frac{t_{k+1} + \lambda t_k}{u_{k+1} + \lambda u_k}; \\ \frac{Y_k}{-x_k} &= \frac{d(u_k + \lambda u_{k+1})}{t_k + \lambda t_{k+1}}, & \frac{-Y_{-1-k}}{-x_{-1-k}} &= \frac{d(u_{k+1} + \lambda u_k)}{t_{k+1} + \lambda t_k} \end{aligned}$$

$$(k = 0, 1, 2, 3, \cdots).$$

It is evident from (53), (54), and (62) that an x- or y-ambiguous set is characterized by having $\lambda = 0$ or 1 in (76) or (77) respectively. Write $\epsilon = 1$ or -1 according as $an > 0$ or $an < 0$. Then as (x, y) runs through one half the elements of an x-ambiguous set (the other half consisting of the values $(-x, -y)$) the ratio $(2ax+by)/y$ assumes precisely once each of the values

(78) $$\frac{t_{2h+\lambda} + 2\epsilon}{u_{2h+\lambda}} \qquad (h = 0, \pm 1, \pm 2, \cdots),$$

λ having a fixed value 0 or 1 for the set. (Corresponding to $\lambda = h = 0$ this relation is to be interpreted as $2ax+by = 0$ if $\epsilon = -1$ and as $y = 0$ if $\epsilon = 1$.) For a y-ambiguous set we interchange a and c, x and y in the preceding. In view of $(2ax+by)^2 - dy^2 = 4an$ we have (47), and similarly (48).

It follows that for an x-ambiguous set one of $4an$, $-4and$, $(t_1+2)an$, or $-(t_1-2)an$ is a square; and similarly for a y-ambiguous set with a replaced by c.

A glance at (76) (where now $0 < \lambda < 1$ or $\lambda > 1$) demonstrates the following theorem.*

* A unification of two types of interval is obtained by means of (62).

THEOREM 8. *Let the discriminant d of the primitive integral form $[a, b, c]$ be positive but not square. If a set of solutions of* (45) *is not x-ambiguous, it contains precisely one pair (x, y) and $(-x, -y)$ satisfying*

$$(79) \qquad y \neq 0 \text{ and } \left|\frac{2ax + by}{y}\right| > \frac{t_1 + 2}{u_1} \text{ if } an > 0,$$

$$(80) \qquad 0 < \left|\frac{2ax + by}{y}\right| < \frac{t_1 - 2}{u_1} \text{ if } an < 0.$$

More generally, let k denote any integer $\geqq 0$. In any set which is not x-ambiguous occurs just one (x, y) and $(-x, -y)$ satisfying

$$(81) \qquad \frac{t_k + 2}{u_k} > \left|\frac{2ax + by}{y}\right| > \frac{t_{k+1} + 2}{u_{k+1}} \text{ if } an > 0,$$

$$(82) \qquad \frac{t_k - 2}{u_k} < \left|\frac{2ax + by}{y}\right| < \frac{t_{k+1} - 2}{u_{k+1}} \text{ if } an < 0.$$

We may designate as Theorem 8′ the analogous result for non-y-ambiguous sets, obtained by interchanging x and y, a and c throughout Theorem 8.

Every solution is contained within one of these intervals.

By (21) and (45) the preceding systems of inequalities may be replaced by the following, k denoting any integer $\geqq 0$:

$$(83) \qquad \begin{aligned} an(t_k - 2) &< dy^2 < an(t_{k+1} - 2) \text{ if } an > 0, \\ -an(t_k + 2) &< dy^2 < -an(t_{k+1} + 2) \text{ if } an < 0, \end{aligned}$$

if the set is not x-ambiguous. In every such set there is precisely one (x, y) and $(-x, -y)$ within each of the intervals (83) for $k = 0, 1, 2, \cdots$; for a non-y-ambiguous set the corresponding intervals are

$$(83') \qquad \begin{aligned} cn(t_k - 2) &< dx^2 < cn(t_{k+1} - 2) \text{ if } cn > 0, \\ -cn(t_k + 2) &< dx^2 < -cn(t_{k+1} + 2) \text{ if } cn < 0. \end{aligned}$$

It may be noted that, if $b = 0$, then y satisfies (83) if and only if x satisfies (83′) for the same k.

THEOREM 9. *Let the discriminant d of the primitive integral form $[a, b, c]$ be positive but not square. Let (t_1, u_1) be the least positive solution of $t^2 - du^2 = 4$. Then the number of sets of solutions of $ax^2 + bxy + cy^2 = n$ is equal to the number of solutions (x, y) of this equation satisfying*

$$(84) \qquad 2|an| - 2an \leqq dy^2 \leqq t_1|an| - 2an, \; y \geqq 0,$$

with the convention that solutions with y at an end point of this interval are counted as $\frac{1}{2}$.

In place of (84) we may use

$$(84') \qquad 2|cn| - 2cn \leqq dx^2 \leqq t_1|cn| - 2cn, \ x \geqq 0;$$

or indeed any of the infinitely many intervals (79)-(83′), with $\leqq$ in place of $<$, and with the same convention for end points.

4.1. **Reduction formulas for primes p not dividing d.** If $(d|p)=-1$, (56) requires $x\equiv y\equiv 0 \pmod p$, so that (10) is obvious.

Hence let $(d|p)=1$ and employ the notations and hypotheses of §3.3. Then (56) requires

$$(85) \qquad x = m_iy + pX,$$

X integral ($i=1$ or 2). The equation (85) defines a (1, 1) correspondence between the integral solutions (x, y) of (56) satisfying $x\equiv m_iy \pmod p$ and the integral solutions (X, y) of

$$(86) \qquad n = apX^2 + (2am_i + b)Xy + p^{-1}(am_i^2 + bm_i + c)y^2.$$

Here $g_i=[ap, 2am_i+b, \cdots]$ is a primitive integral form of discriminant d. The solutions (X, y) of (86) in which $p|y$ correspond to solutions (X, Y) of

$$(87) \qquad n/p = aX^2 + (2am_i + b)XY + (am_i^2 + bm_i + c)Y^2,$$

under the transformation $y=pY$. The form in (87) is equivalent to $f=[a, b, c]$. Hence to conclude for every integer n that

$$f(pn) = f(n/p) + \{g_1(n) - f(n/p)\} + \{g_2(n) - f(n/p)\},$$

that is,

$$(88) \qquad f(pn) + f(n/p) = g_1(n) + g_2(n),$$

we have to prove that if (x, y) and (x_0, y_0) are equivalent solutions of (56), then (X, y) and (X_0, y_0), where

$$(89) \qquad x = m_iy + pX, \quad x_0 = m_iy_0 + pX_0,$$

are equivalent solutions of (86), and conversely. The converse holds by Theorem 6. Assume that (20) holds. Then by (89),

$$X = p^{-1}(x - m_iy) = \tfrac{1}{2}[t - (2am_i + b)u]X_0 - p^{-1}(am_i^2 + bm_i + c)uy_0,$$
$$y = apuX_0 + \tfrac{1}{2}y_0[t + (2am_i + b)u], \text{ as required.}$$

The remaining developments of §6, MZ, may now be carried through, if we use Theorems 4 and 5.

4.2. **Reduction formulas for primes dividing d.** First, let $p>2$, $d=b^2-4ac$, $p|d$, where a, b, c are relative-prime integers. Then p does not divide a or c, say p does not divide a. We can choose integers A and Q such that

$$(90) \qquad 2Aa + Qp = 1.$$

Then the equation

$$(91)\qquad ax^2 + bxy + cy^2 = pn,$$

being equivalent to $(2ax+by)^2 - dy^2 = 4apn$, implies

$$(92)\qquad x = -Aby + pX, \quad X \text{ integral}.$$

Second, let $p=2$, $4\,|\,d$. Then a or c is odd, say a. Then (91) implies $x \equiv cy$ (mod 2), so that in place of (92) we have

$$(93)\qquad x = 2X \text{ if } c \text{ is even}, \; x = 2X + y \text{ if } c \text{ is odd}.$$

Thus (92) or (93) sets up a (1, 1) correspondence between the integral solutions (x, y) of (91) and the integral solutions (X, y) of

$$(94)\qquad apX^2 + b'Xy + c'y^2 = n,$$

where, in the respective cases (92), (93_1), and (93_2),

$$(95)\qquad \begin{array}{lll} b' = bQp, & c' = (aA^2b^2 - Ab^2 + c)/p & (p > 2); \\ b' = b, & c' = \frac{1}{2}c & (p = 2, c \text{ even}); \\ b' = 2a + b, & c' = \frac{1}{2}(a + b + c) & (p = 2, c \text{ odd}). \end{array}$$

It is plain that c' is an integer in all cases, and that the form

$$(96)\qquad g = [ap, b', c']$$

is of discriminant d. Hence, in case (95_1), $p\,|\,c'$ if and only if $p^2\,|\,d$, so that g is primitive if and only if p^2 does not divide d. In either of cases (95_2) or (95_3), the divisor of g is seen to be 1 if $d \equiv 8$ or 12, but 2 if $d \equiv 0$ or 4 (mod 16).

In the respective three cases, write

$$(97)\qquad \begin{array}{l} X = (x + Aby)/p, \;\; X_0 = (x_0 + Aby_0)/p; \;\; X = \frac{1}{2}x, \;\; X_0 = \frac{1}{2}x_0; \\ X = (x - y)/2, \qquad X_0 = (x_0 - y_0)/2. \end{array}$$

We find, by (90) and the values of b' and c', that the relations (20) are equivalent to

$$(98)\qquad \begin{array}{l} X = \frac{1}{2}(t - b'u)X_0 - c'uy_0, \\ y = apuX_0 + \frac{1}{2}(t + b'u)y_0. \end{array}$$

Thus, *if g is primitive*, that is, if (6) does not hold, the (1, 1) correspondence between representations set up by (92) or (93) carries over to sets of representations. Hence, if $f = [a, b, c]$,

$$(99)\qquad f(pn) = g(n).$$

4.3. Finally let (6) hold. Now (92) or (93) still sets up a (1, 1) correspondence between the solutions (x, y) of (91) and the solutions (X, y) of

$$aX^2 + (b'/p)Xy + (c'/p)y^2 = n/p. \tag{100}$$

But now, while $(x, y)\sim(x_0, y_0)$ implies $(X, y)\sim(X_0, y_0)$, the converse is no longer true. The former fact is evident from (98) on replacing ap by a, b' by b'/p, u by pu, c' by c'/p, and on noticing that $t^2-(d/p^2)(pu)^2=4$.

In the cases where $T=\pm 2$, $U=0$ are the only solutions of

$$T^2 - (d/p^2)U^2 = 4, \tag{101}$$

we evidently have conversely that

$$(X, y) \sim (X_0, y_0) \text{ implies } (x, y) \sim (x_0, y_0), \tag{102}$$

the first equivalence relating to (100), the second to (91). If $d=-3p^2$ or $-4p^2$, the values of σ in (14) are evident from the relative numbers of solutions of (20) and (101) (which are the numbers of solutions in a set).

There remains only the case where d is positive but not square. Then $(X, y)\sim(X_0, y_0)$ may be written as

$$\begin{aligned} 2aX + (b'/p)y &= \tfrac{1}{2}T(2aX_0 + (b'/p)y_0) + \tfrac{1}{2}(d/p^2)Uy_0, \\ y &= \tfrac{1}{2}U(2aX_0 + (b'/p)y_0) + \tfrac{1}{2}Ty_0, \end{aligned} \tag{103}$$

where (T, U) denotes some solution of (101).

Our remaining problem may now be stated precisely. We are given a pair of integers x_0, y_0 such that X_0, as defined in (97), is an integer. Let K denote the aggregate of all (integer) pairs x, y defined by (103), and (97), as T, U range over all the solutions of (101). Each such pair x, y is a solution of (91). Evidently K is the sum of a certain number of sets of such solutions. The problem is to determine that number.

Let then x', y' be another pair, defined by

$$\begin{aligned} 2aX' + (b'/p)y' &= \tfrac{1}{2}T'(2aX_0 + (b'/p)y_0) + \tfrac{1}{2}(d/p^2)U'y_0, \\ y' &= \tfrac{1}{2}U'(2aX_0 + (b'/p)y_0) + \tfrac{1}{2}T'y_0, \end{aligned} \tag{104}$$

and by

$$X' = (x' + Aby')/p, \quad X' = \tfrac{1}{2}x', \quad X' = (x' - y')/2, \tag{105}$$

respectively; T', U' being another solution of (101). Changing (x, y) to $(-x, -y)$ or (x', y') to $(-x', -y')$ if necessary, we may suppose $(T, U)=(T_k, U_k)$ and $(T', U')=(T_l, U_l)$, where T_k, U_k play the same role for (101) as t_k, u_k for $t^2-du^2=4$.

Let σ denote the least positive index such that $U_\sigma\equiv 0 \pmod p$. Hence $u_n=U_{n\sigma}/p$ and $t_n=T_{n\sigma}$ $(n=0, \pm 1, \pm 2, \cdots)$.

On solving for $2aX_0+(b'/p)y_0$ and y_0 from (104) and substituting in (103), and using relations for T_k, U_k analogous to (58) and (59_1), we obtain

$$
\begin{aligned}
2aX + (b'/p)y &= \tfrac{1}{2}T_{k-l}(2aX' + (b'/p)y') + \tfrac{1}{2}(d/p^2)U_{k-l}y', \\
y &= \tfrac{1}{2}U_{k-l}(2aX' + (b'/p)y') + \tfrac{1}{2}T_{k-l}y'.
\end{aligned}
\tag{106}
$$

Now, in all three cases $p[2aX+(b'/p)y]=2ax+by$. Hence (106) may be written

$$
\begin{aligned}
2ax + by &= \tfrac{1}{2}T_{k-l}(2ax' + by') + \tfrac{1}{2}d(U_{k-l}/p)y', \\
y &= \tfrac{1}{2}(U_{k-l}/p)(2ax' + by') + \tfrac{1}{2}T_{k-l}y'.
\end{aligned}
\tag{107}
$$

Now since $(2ax'+by')^2-dy'^2\neq 0$, these equations cannot hold with T_{k-l} and U_{k-l}/p replaced by any other numbers. Hence we see that (x, y) and (x', y') defined by (103) and (104) with $(T, U)=(T_k, U_k)$ and $(T', U')=(T_l, U_l)$ are equivalent if and only if $k\equiv l \pmod{\sigma}$. This proves (13) with σ as in (16).

McGill University,
Montreal, Canada

ON ANALYTICAL COMPLEXES*

BY

S. LEFSCHETZ AND J. H. C. WHITEHEAD

1. In his Colloquium Lectures† one of us outlined a proof of an important theorem regarding the covering of analytic loci by complexes. A proof for algebraic varieties had previously been given by B. van der Waerden‡ and B. O. Koopman and A. B. Brown§ have recently proved the theorem for analytic loci. The object of this paper is to give a detailed proof along the lines indicated in *Topology*.

2. We begin with certain general observations|| concerning the nature of a configuration ξ (at first complex) represented by an analytic system

$$(2.1) \qquad F_h(x_1, \cdots, x_n) \equiv F_h(x) = 0 \qquad (h = 1, 2, \cdots, r),$$

in the vicinity of a given point O of ξ which we take as the origin throughout for the complex euclidean space S_n containing ξ. There is a neighborhood of O relative to ξ consisting of a finite number of algebroid elements, any one of them, say w_p, having about its center O, in a suitable coordinate system y_i, a canonical representation

$$(2.2) \quad \begin{aligned} &(\mathrm{a}) \qquad H(y_1, \cdots, y_p, y_{p+1}) = 0, \\ &(\mathrm{b}) \qquad \frac{\partial H}{\partial y_{p+1}} \cdot y_{p+1+i} + G_i(y_1, \cdots, y_{p+1}) = 0, \end{aligned}$$

where H, G_i are *pseudopolynomials* in y_{p+1}, i.e. polynomials with coefficients analytic in $y_1, \cdots, y_p$ at $(y) = (0)$, and where moreover H is algebraically irreducible and *special*, i.e. its leading coefficient is unity and its other coefficients are zero at (0). p is the *complex dimension* of w_p (dim w_p), and also of ξ at O ($\dim_O \xi$) when dim $w = p$ for some w component of ξ at O, and $\leqq p$ for all others. When O is not on ξ we agree to take $\dim_O \xi = -1$.

We have the following basic *irreducibility property*: if ξ does not contain w_p, then the intersection $\xi \cdot w_p$ is a ξ, whose dimension $< p$ at O. For the case

---

* Presented to the Society, August 31, 1932; received by the editors in July, 1932, and (revised) September 22, 1932.

† S. Lefschetz, *Topology*, Colloquium Publications, vol. XII, New York, 1930, p. 364. Except as introduced here the same notation and terminology will be used as in *Topology*.

‡ Mathematische Annalen, vol. 102, pp. 337–362.

§ These Transactions, vol. 34 (1932), pp. 231–252.

|| Based on Osgood's *Lehrbuch der Funktionentheorie*, vol. II, chapter II.

where ξ is defined by a single relation (2.1) see Osgood's proof (loc. cit., p. 133), and the extension to any ξ is obvious.

We shall now recall a series of properties most of them direct consequences of the preceding.

I. The solution of an infinite system (2.1) about any point O is of the same type as for a finite system.

II. A point of w_p is *singular* if the rank of the Jacobian matrix J of (2.2) is $<n-p$ at the point; it is an *ordinary* point otherwise. The locus α of the singular points is the *singular* locus of w_p. Since J contains a minor of order $n-p$ equal to $(\partial H/\partial y_{p+1})^{n-p}\not\equiv 0$ when $H=0$, the conditions that the rank be $<n-p$ define a ξ not containing w_p. Hence $\alpha\cdot w_p=\alpha$ is a ξ and $\dim_O \alpha<p$.

The characteristic property of an ordinary point (a) is to have relative to w_p a neighborhood which is a $2p$-cell E_{2p} with a parametric representation

$$x_i - a_i = \phi_i(u_1, \cdots, u_p), \tag{2.3}$$

where at $(u)=(0)$ the ϕ's are analytic, vanish and have a Jacobian matrix of rank p. Every point of w_p is a limit-point of ordinary points.

III. It is impossible to decompose w_p about O into a sum of $r>1$ sets $w^i_{q_i}$. For otherwise $w^i=w^i\cdot w_p\neq w_p$, hence $q_i<p$. Therefore w_p would have points about which the coordinates depend upon $q_i<p$ parameters, which is untrue. As a noteworthy consequence the resolution of ξ into w components about O is unique and hence $\dim_O \xi$ depends solely upon O and ξ.

IV. Given a fixed coordinate system x_i we shall call *vertical* the direction of its x_n axis and denote by $P(\lambda)$ the projection of the locus λ on $x_n=0$. If the center O of w_p is an isolated intersection with the vertical through O, then $P(w_p)$ is a w_p of center $P(O)$. This does not require that the coordinates x_i be canonical for w_p. We may of course assume that O is the origin so that $P(O)=O$. Under the assumption w_p may be represented by a system (2.1) such that no $F_h(0, \cdots, 0, x_n)\equiv 0$, hence we may replace all the F's by pseudopolynomials in x_n. The algebraic elimination of x_n yields then a system such as (2.1) without x_n, representing $P(w_p)$; hence $P(w_p)$ is a ξ. If this ξ had $r>1$ components about O the vertical cylinders erected on them would decompose w_p into a ξ having at least r components w about O. Therefore $r=1$ and $P(w_p)$ is a w of center O. If a point Q varies on w_p, $x_n(Q)$ is a finite-valued function of $P(Q)$, hence $P(Q)$ depends on p parameters and $P(w_p)$ is a w_p.

Since x_n is a finite-valued function on $P(w_p)$ we have for w_p a representation (Osgood, loc. cit., p. 114)

$$\text{(2.4)}\quad \begin{aligned} &\text{(a)} \qquad G_i(x_1, \cdots, x_{n-1}) = 0,\\ &\text{(b)} \qquad H(x_1, \cdots, x_n) = 0, \end{aligned}$$

where H is a pseudopolynomial in x_n and (2.4a) represents $P(w_p)$ in $x_n=0$. Since no true subset of w_p is a w_p, H is irreducible.

The branch locus β of w_p is its intersection with $\partial H/\partial x_n=0$. Just as for the singular locus we have $\dim_O\beta<p$. Hence w_p possesses ordinary points not on β.

3. We shall now consider a real analytic variety η. It is a real locus represented by a real system (2.1) or system with F's all real.* The same system represents a ξ to be denoted by (η). Let O be a point of η. On following up Osgood's resolution of (η) into w components about O we find that their canonical coordinates y_i may be chosen real. This being assumed done we have for a component w_p three possibilities: (a) The canonical system of w_p is real and w_p possesses real ordinary points. The real subset of w_p (real algebroid element) will be denoted by v_p, so that $w_p=(v_p)$; incidentally the form of (2.2) shows that when $p=0$, O is an ordinary point, i.e., it is a v_0. (b) This case is the same as the preceding except that the real points of w_p are all singular. (c) The canonical system of w_p cannot be chosen real. When $w_p=w_p$, H and G_i in (2.2) may be replaced by $H+\overline{H}$, $G_i+\overline{G}_i$, both real and of the same form, hence we have cases (a) or (b). Therefore in case (c) necessarily $w_p\neq\widehat{w}_p$.

We shall now show that η may be decomposed about O into a finite sum of v's. Let $p=\dim_O(\eta)$. Since the required result holds when $p=0$ we use induction on p. The real points of a w_p of type (b) are on the singular locus of w_p which is an (η) whose dimension at O is $<p$. As regards the real points of a w_p of type (c), let $f_i=0$ be the canonical equations of w_p. Since (2.1) is real, $\bar{f}_i=0$ are the canonical equations of another component of (η) which is $\widehat{w}_p$. Hence the real points in question are on $w_p\cdot\widehat{w}_p$ and since w_p does not contain $\widehat{w}_p$, this is a ξ whose dimension at O is $<p$. But this ξ, being represented by the real system $f_i+\bar{f}_i=0$, $-i(f_i-\bar{f}_i)=0$, is also an (η). The real points of components not of type (a) being thus on varieties (η) whose dimensions at O are $<p$, the required result is a consequence of the hypothesis of the induction.

The meaning of $\dim v_p$, $\dim_O\eta$ is as before. As it happens they are precisely the Urysohn-Menger dimensions, but this does not matter for our purpose.

The *irreducibility property* holds for η: if η does not contain v_p, $\dim_O\eta\cdot v_p<p$. Its proof is as follows. Under the hypothesis (η) does not contain (v_p), hence $p>\dim_O(\eta\cdot v_p)=\dim_O\eta\cdot v_p$.

Properties I, $\cdots$, IV hold with v in place of w and with these modifications: (a) (2.3) represents a real analytic E_p; (b) (2.4) still represents v_p in the

* The condition $\overline{F}(\bar{x})=\overline{F(x)}$ defines an analytic function $\overline{F}$, the *conjugate* of F, and F is *real* whenever $\overline{F}\equiv F$. The set of the conjugate points of the points of a locus λ will be denoted by $\widehat{\lambda}$, the usual "bar" notation being reserved for the closure.

real and (v_p) in the complex domains, but (2.4a) represents a real v_p' which may be $\neq P(v_p)$, since in addition it may contain points which are the projections of pairs of conjugate points of (v_p). Thus we can only assert that $P(v_p)$ is a subset of a v_p'. Here again v_p contains an ordinary point Q not on the branch locus β. Q possesses then relative to v_p a neighborhood which is an analytic E_p homeomorphic with $P(E_p)$. This implies that in (2.3) the Jacobian matrix of $\phi_1, \cdots, \phi_{n-1}$ is of rank p at $(u)=(0)$. Hence $P(Q)$ has a neighborhood relative to $P(v_p)$, and not merely relative to v_p', which is an analytic E_p. We may think of $P(Q)$ as an ordinary point of $P(v_p)$.

Henceforth we shall deal exclusively with the real domain.

4. **The segments on v_p.** Let α_h be direction cosines for S_n, so that (α) is a point of the unit-sphere H_{n-1} of S_n. A point (x) of v_p $(p<n)$ will not be an isolated intersection of v_p with the line $x_h+s\alpha_h$ (s variable) when and only when the MacLaurin series for s of the functions $f_i(x+s\alpha)$ are $\equiv 0$, where the f's are the left-hand sides of a representation (2.1) for v_p. There results a real analytic system

$$\Phi_j(x;\alpha) = 0. \tag{4.1}$$

Its solutions for $(x;\alpha)$ in the vicinity of any solution $(x^0;\alpha^0)$ make up a finite number of sets v_q. On such a v_q we shall then have a parametric representation

$$\text{(a)}\quad x_i = \phi_i(y_1, \cdots, y_q);\quad \text{(b)}\quad \alpha_i = \psi_i(y_1, \cdots, y_q), \tag{4.2}$$

where ϕ_i, ψ_i are analytic on v_q. The system (4.2b) represents on H_{n-1} the directions near (α^0) corresponding to segments on our given v_p associated with (4.2).

Since (4.2) represents a v_q,

$$\left\| \frac{\partial\phi_i}{\partial y_j};\ \frac{\partial\psi_i}{\partial y_j} \right\| \tag{4.3}$$

is of rank q at some points as near as we please to (y^0). On the other hand, for $y_1, \cdots, y_q$ near $y_1^0, \cdots, y_q^0$ and s arbitrary but small, $\phi_i+s\psi_i$ represents a point of our initial v_p, and hence among these functions at most p are functionally independent, or

$$\left\| \frac{\partial\phi_i}{\partial y_j} + s\frac{\partial\psi_i}{\partial y_j};\ \psi_i \right\| \tag{4.4}$$

is of rank $\leqq p$, and this must hold for s small but arbitrary. Now any determinant of this matrix containing s is a polynomial in s whose leading coefficient is the corresponding determinant of

$$(4.5) \qquad \left\| \frac{\partial \psi_i}{\partial y_j}; \psi_i \right\|$$

which must therefore be of rank $\leqq p$. Owing to the relations

$$\sum \psi_i^2 = 1, \quad \sum \psi_i \frac{\partial \psi_i}{\partial y_j} = 0,$$

the new matrix may be bordered with a row $0, \cdots, 0, 1$ without changing its rank. It follows that the rank of

$$\left\| \frac{\partial \psi_i}{\partial y_j} \right\|$$

is at most $p-1 \leqq n-2$. Therefore the directions of segments meeting v_p in an infinite set are represented on H_{n-1} by a variety η whose dimension at any point $<n-1$, and hence they are nowhere dense on the sphere.†

5. **Analytic complexes.** By an *analytic structure* ζ we shall mean a real point set in a real S_n which constitutes a topological space with varieties η as the neighborhoods. Each point Q of ζ has then a neighborhood relative to ζ made up of a finite set $v_{q_1}{}^1, \cdots, v_{q_r}{}^r$, where the v's have Q as their common center. The largest q_r is the *dimension of* ζ *at* Q ($\dim_Q \zeta$), and the largest value p of $\dim_Q \zeta$ for Q on ζ is the *dimension* of ζ ($\dim \zeta$) which is then designated by ζ_p.

We now define the point Q of ζ, whose neighborhood is $v_{q_1}{}^1 + \cdots + v_{q_r}{}^r$, as *singular* when $r>1$, or when $r=1$ and $\dim_Q \zeta < p$, or else $\dim_Q \zeta = p$ and Q is singular for its unique v. From property II of §2 for an η, we have that the set of all singular points or *singular locus* is a ζ_r, $r<p$. A point of $\zeta_p - \zeta_r$ is an *ordinary* point of ζ_p. Its characteristic property is that it possesses relative to ζ_p a neighborhood which is an analytic E_p.

By an *analytic p-element*, or merely p-element, ϵ_p, we shall mean a relatively open subset of a structure ζ_p, containing at least one ordinary point, and such that $\bar{\epsilon}_p \subset \zeta_p$. Under these conditions we shall describe ϵ_p and ζ_p as *associated* with each other. By an *analytic p-complex*, κ_p, we shall mean a finite set of non-intersecting elements ϵ, of dimension up to and including p, which constitute a closed bounded point set in S_n. By convention the empty set is to be a ζ_{-1}, an ϵ_{-1} or a κ_{-1}. We shall write $F(\zeta) = \bar{\zeta} - \zeta$. We do not consider here infinite κ's, since they may be taken care of as in *Topology*.

The intersection of two or more ζ's or ϵ's is respectively a ζ or an ϵ. If $\kappa = \sum \epsilon$, $\kappa^* = \sum \epsilon^*$ are complexes, so is $\kappa \cdot \kappa^* = \sum \epsilon \cdot \epsilon^*$. Similarly when $\kappa \cdot \zeta$ is closed then $\kappa \cdot \zeta$ is the complex $\sum \epsilon \cdot \zeta$.

† Cf. Koopman and Brown, loc. cit., p. 242.

A complex κ' will be called a *subdivision* of κ if the two coincide as point sets and if each element of κ' is contained in one of κ. If κ^* is any complex on κ it is clear from the preceding paragraph that κ has a subdivision with one of κ^* as a subcomplex. It follows that $\kappa+\kappa^*$ can be covered by a complex having subdivisions of κ and κ^* as subcomplexes. For κ and κ^* can each be subdivided to form a complex having a common subcomplex covering $\kappa\cdot\kappa.^*$

Whenever throughout κ we have $\epsilon_p\cdot\bar{\epsilon}_q=0$ for $q\leqq p$, κ is said to be *normal*. When κ is normal, it remains closed, and hence a κ (moreover a normal κ) when one or more p-elements are removed from it.

Every complex has a normal subdivision. Given any ζ_p, we shall denote its singular locus by $\zeta_{p'}$ $(p'<p)$. Let then ϵ_p, ϵ_q be two elements of κ_p and ζ_p, ζ_q associated structures. We shall first show that there exists a $\zeta_r\supset\epsilon_p\cdot\bar{\epsilon}_q$ such that $r<p$ and that the distance from $\epsilon_p\cdot\bar{\epsilon}_q$ to $F(\zeta_r)>0$. In any case $\epsilon_p\cdot\bar{\epsilon}_q\subset\bar{\epsilon}_p\cdot\bar{\epsilon}_q\subset\zeta_p\cdot\zeta_q=\zeta_s$. Also $F(\zeta_s)\subset F(\zeta_p)+F(\zeta_q)$. Since no $\bar{\epsilon}$ meets its $F(\zeta)$, $\bar{\epsilon}_p\cdot\bar{\epsilon}_q$ does not meet $F(\zeta_s)$, and as $\bar{\epsilon}_p\cdot\bar{\epsilon}_q$ is self-compact the distance of the two sets >0. Therefore when $s<p$ we may take $\zeta_r=\zeta_s$.

Let now $s=p$ and let Q be a point of $\epsilon_p\cdot\bar{\epsilon}_q$ not on $\zeta_{s'}$, so that $Q\subset\zeta_s-\zeta_{s'}$ and $\dim_Q\zeta_s=p$. This implies $q=p$ and that the nieghborhoods of Q relative to ζ_p and ζ_q have a common v_p which is then wholly on ϵ_p near Q. In that case necessarily $Q\subset\zeta_{q'}$. For otherwise v_p would be a complete neighborhood of Q relative to ζ_q, hence it would contain points of ϵ_q infinitely near Q, and we should have $\epsilon_p\cdot\epsilon_q\neq 0$, which is ruled out. It follows $\epsilon_p\cdot\bar{\epsilon}_q\subset\zeta_{q'}+\zeta_{s'}$.

Since a singular locus is closed relative to its ζ, and since $F(\zeta_s)\subset F(\zeta_p)+F(\zeta_q)$, we find that $\zeta_r=\zeta_{q'}+\zeta_{s'}-F(\zeta_p)-F(\zeta_q)$ satisfies the condition for a structure, with $F(\zeta_r)\subset F(\zeta_p)+F(\zeta_q)$. Since the last two F's do not meet $\bar{\epsilon}_p\cdot\bar{\epsilon}_q$, this is likewise the case as regards $F(\zeta_r)$, which implies also $\zeta_r\supset\bar{\epsilon}_p\cdot\bar{\epsilon}_q\supset\epsilon_p\cdot\bar{\epsilon}_q$, and that the distance condition holds. Since $r=p'$ or q', both $<p$, ζ_r has all the properties that we require.

We can find a closed polyhedral neighborhood of $\bar{\epsilon}_p\cdot\bar{\epsilon}_q$ not meeting $F(\zeta_r)$, and its intersection with ζ_r is a κ_r. The sum of these complexes for all p-elements is a κ_t, $t<p$, and $\epsilon_p'=\epsilon_p-\kappa_t$ is an element. Replacing ϵ_p by ϵ_p' together with the sum of the elements $\epsilon_p\cdot\kappa_t$, we obtain a subdivision κ_p', such that $\epsilon_p'\cdot\bar{\epsilon}_q'=0$ if $\epsilon_q'\neq\epsilon_p'$. Hence $\kappa_p'-\sum\epsilon_p'$ is a κ whose dimension $<p$. The required result follows then by induction on p.

6. **The covering theorem.** Let κ be any complex and let "vertical" direction or projection have the same meaning as in §2, IV. Every point of κ has a neighborhood relative to κ made up of a finite number of v's. Since κ is self-compact it can be covered with a finite number of v's. It follows then from §4 that the axes may be so chosen that no vertical meets κ in an infinite set.

From §§3, 5, we conclude that κ has a subdivision (obtained as its intersection with a suitable polyhedral complex) whose elements are each on a v represented by a system (2.4). The subdivision can then be normalized so that at each step in the process the preceding property is preserved. Ultimately we turn the complex into a normal complex, still called κ, whose elements all have the property just described.

Let ϵ_p be any element of our new κ with its ζ_p given by (2.4). The branch locus ζ^* is of dimension $<p$. We verify at once that $\epsilon^*=\epsilon_p\cdot\zeta^*$ and $\epsilon_p-\epsilon^*$ are elements with ζ^* and ζ_p as associated structures. Referring to the end of §3 we find also that, when $p<n, P(\epsilon_p-\epsilon^*)$ is an ϵ_p. Moreover when $(x_1, \cdots, x_{n-1})$ ranges over $P(\epsilon_p-\epsilon^*)$, to certain real roots x_n' of $H=0$ there will correspond points $(x_1, \cdots, x_{n-1}, x_n')$ which generate elements ϵ_p' whose sum is $\epsilon_p-\epsilon^*$.

Assuming that our complex is a κ_p, $p<n$, we decompose every ϵ_p of κ_p in the set of p-elements ϵ_p' plus ϵ^* (whose dimension $<p$), and repeat the operation for the elements of next lower dimension of the new complex, etc. Ultimately then we have in place of κ_p a new normal complex, still to be called κ_p, such that every ϵ of κ_p has for projection $P(\epsilon)$ an element, and ϵ is represented by an analytic relation $x_n=f(Q)$, $Q\subset P(\epsilon)$ (analytic homeomorphism).

A final subdivision $\kappa_p'=\sum\epsilon'$ of κ_p will now be made, such that $P(\kappa_p')$ can be covered with a $\kappa_p''=\sum\epsilon''$ having the property that every $P(\epsilon')$ is an exact sum of elements ϵ''. For $p=0$ this is trivial, hence we use induction on p. Taking κ_p in the reduced form just obtained, $\kappa_q=\kappa_p-\sum\epsilon_p$ is a complex with $q<p$. Under the hypothesis of the induction it has a subdivision $\kappa_q'=\sum\epsilon'$ of the desired type. Let δ be a positive number such that $-\delta<x_n<\delta$ on κ_p, and let $C(\kappa_q')$ be the $(q+1)$-complex whose elements are the parts of the vertical cylinders based on the ϵ''s lying between the spaces $x_n=\pm\delta$, together with their intersections with these spaces. Let ϵ_r be an element of the original κ_p. Since an ϵ_r carries no vertical segment, the intersection $\epsilon_r\cdot C(\kappa_q')$ consists of elements of dimension $\leqq q$, some being of dimension q when $r=q$. Therefore $\kappa_p\cdot C(\kappa_q')$ is a q-complex, and since $q<p$, it has a subdivision $\kappa_q^{*\prime}$ such that $P(\kappa_q^{*\prime})$ is covered by a κ_q'' of the required type. Given any ϵ_p of κ_p we form a new element $\epsilon_p'=\epsilon_p-\kappa_q^{*\prime}$. Then $\kappa_p'=\kappa_q^{*\prime}+\sum\epsilon_p'$ is the required subdivision of κ_p. For let κ_p' contain m p-elements $\epsilon_p'^{\alpha}$ and let $\eta^\alpha=P(\epsilon_p'^{\alpha})$. When $m=1$, we can take $\kappa_p''=\kappa_q''+\eta'$. Therefore we may use induction on m. Removing ϵ'^m from κ_p' we have a complex $\kappa_p^{*\prime}$ which, under the hypothesis of the induction, possesses an associated $\kappa_p^{*\prime\prime}$ covering $P(\kappa_p^{*\prime})$. Now $\eta'=\eta^m-\kappa_p^{*\prime\prime}$ is also an element and $\kappa_p''=\eta'+\kappa_p^{*\prime\prime}$ is a covering of $P(\kappa_p')$ such as we are seeking.

Observe that every ϵ' is still analytically homeomorphic with its projec-

tion $P(\epsilon')$ since this holds as regards the ϵ of κ_p on which it lies. We are now ready for the

THEOREM. *Every analytic complex has a simplicial subdivision.*

We first assume $p<n$, and κ_p in its ultimate reduced form, κ_p', κ_p'' having the same meaning as above. The theorem being trivial for $n=0$ we use induction on n. Since κ_p'' is on an S_{n-1} it has then a simplicial subdivision $K_p^0=\sum\sigma$. Let QQ' be any vertical with $Q\subset\sigma_q$. It intersects κ_p' in points $Q^1, Q^2, \cdots, Q^r$ (r finite) each on a different ϵ' of κ_p', say $Q^i\subset\epsilon'^i$. When Q ranges continuously over σ_q, by the above Q^i remains on ϵ'^i and generates a homeomorph of σ_q, a cell $E_q^i\subset\epsilon'^i$, and no two of these cells intersect. As a consequence $K_p=\sum E_q^i$ is a cellular subdivision of κ_p' and hence of κ_p. We shall now show that the homeomorphism between E_q and E_q^i can be extended to their boundaries. This merely requires that we prove that when $Q\subset F(E_q)$, it has a unique image on $F(E_q^i)$. Suppose that it has s images Q'^j. We may choose for each Q'^j a neighborhood relative to $E_q{}^i$ consisting of a cell $E_q{}^{ij}$ whose projection is a simplex σ_q^j (in the straightness of σ_q), no two of the cells $E_q{}^{ij}$ intersecting. As a consequence $\sigma^j\cdot\sigma^h=0$ $(j\neq h)$ and Q has for neighborhood relative to σ_q a set of s non-intersecting q-simplexes, which can only be if $s=1$, as asserted. Since $\overline{E}_q{}^i$ and $\bar{\sigma}_q$ are homeomorphic, E^i is simplicial and so is K_p.

If we have a κ_n, on removing its n-elements we have a κ_p, $p<n$, which we identify with the κ_p just considered. When $Q\subset\sigma_q$, the points of κ_n projected on Q may include some of the segments Q^iQ^{i+1} and we observe that, since r is fixed throughout any σ, if the segment is zero anywhere on a face of σ_q it is zero throughout that face. As a consequence we find by an elementary induction that when Q ranges over σ_q the segments $\neq 0$ generate $(q+1)$-cells E_{q+1}^i whose structure is that of a truncated simplicial prism. Since these cells are convex, the covering K_n thus obtained for κ_n is convex, and its first derived, which is simplicial, answers the question.

COROLLARY. *If $\kappa_q\subset\kappa_p$, κ_p has a simplicial subdivision with a subcomplex covering κ_q.*

For κ_p has a subdivision κ_p' having a subdivision of κ_q as a subcomplex. In particular κ_p may be a closed polyhedral region of S_n containing κ_q. This is substantially the theorem of *Topology*, p. 364.

PRINCETON UNIVERSITY,
PRINCETON, N. J.

NON-CONJUGATE OSCULATING QUADRICS OF A CURVE ON A SURFACE*

BY

R. C. BULLOCK

1. **Introduction.** This paper is concerned with making a study of the projective differential geometry of a non-conjugate net of curves on a surface in three-space by means of a pair of osculating quadrics defined in the following manner. Consider a curve C on the surface S. At a point P of C and at two neighboring points P_1, P_2 on C construct the tangents of the curves of one family of the non-conjugate net. The limit of the quadric surface determined by these three lines as the points P_1, P_2 approach P along C is a *non-conjugate osculating quadric* at the point P on C. The other osculating quadric is obtained in a similar manner by drawing tangents to the other family of the net. Thus we associate with each point of the surface a pair of osculating quadrics analogous to the *asymptotic osculating quadrics*† of Bompiani and Klobouček and the *conjugate osculating quadrics*‡ of Lane.

We compute the equations of the osculating quadrics and note some results that follow rather immediately therefrom. The chief contribution of the paper is a complete discussion of the nature of the curve of intersection of the two osculating quadrics at a point of a curve on the surface. We also study the curve of intersection of corresponding osculating quadrics for two curves which are respectively members of two families of curves which form a conjugate net on the surface. As a by-product we obtain a new necessary and sufficient condition that a net of curves on the surface be a conjugate net.

2. **Analytic basis.** In this section we set up an analytic basis for the study of a non-conjugate net of curves on a surface in projective three-space following Green's method.§ We shall also list in this section certain of Green's results that we shall need for later reference. Let it be noted here that we shall assume that the surface sustaining the net is not developable, and that unless otherwise specifically stated the net in question is not the asymptotic net.

* Presented to the Society, December 29, 1932; received by the editors August 25, 1932.

† E. P. Lane, *Projective Differential Geometry of Curves and Surfaces*, University of Chicago Press, 1932, p. 80.

‡ E. P. Lane, *Conjugate nets and the lines of curvature*, American Journal of Mathematics, vol. 53 (1931), p. 577.

§ G. M. Green, *Nets of space curves*, these Transactions, vol. 21 (1920), pp. 207–236.

Let the surface under consideration be the analytic surface whose vector equation in homogeneous coordinates is

$$x = x(u, v).$$

A necessary and sufficient condition that the parametric curves $u=\text{const.}$, $v=\text{const.}$ shall not form a conjugate net is that x shall not satisfy an equation of Laplace of the form

$$ax_{uv} + bx_u + cx_v + dx = 0.$$

This is equivalent to saying that the fourth-order determinant

$$W = (x_{uv}, x_u, x_v, x)$$

does not vanish. Hence the functions x are solutions of a pair of partial differential equations of the form

$$x_{uu} = px + \alpha x_u + \beta x_v + Lx_{uv}, \qquad x_{vv} = qx + \gamma x_u + \delta x_v + Nx_{uv}. \tag{1}$$

If we adjoin certain integrability conditions, which we shall not write down, these equations will form a completely integrable system; that is, any fundamental set of solutions can be expressed as a linear combination, with constant coefficients, of any other fundamental system. We shall base our projective theory of surfaces on this completely integrable system of partial differential equations.

If we regard the u-tangent at a point P_x as a generator of the congruence of u-tangents, its other focal point is given by

$$\rho = x_u + (\beta/L)x. \tag{2}$$

Similarly the focal point of the v-tangent at P_x is given by

$$\sigma = x_v + (\gamma/N)x. \tag{3}$$

The line joining ρ and σ is called by Green the *ray of the point* P_x.

Since the parametric net is not asymptotic the osculating planes of the u-curve and v-curve through P_x determine a line that passes through P_x but does not lie in the tangent plane. This line is called by Green the *axis of the point* P_x. The totality of axes of points of the surface generate the axis congruence. The point τ which is the harmonic conjugate of the point x with respect to the two focal points of the axis is given by

$$\tau = K'x + (\gamma/N)x_u + (\beta/L)x_v + x_{uv}, \tag{4}$$

where the value of K' is given by

$$\begin{aligned} K' = &-\tfrac{1}{2}\big[b^{(21)} + c^{(12)} - (\beta/L)a^{(12)} - (\gamma/N)a^{(21)} + (\alpha\gamma)/N + (\beta\delta)/L \\ &- L\gamma^2/N^2 - N\beta^2/L^2 - 2\beta\gamma/(LN) + (N\gamma_u - N_u\gamma)/N^2 \\ &+ (L\beta_v - L_v\beta)/L^2\big], \end{aligned} \tag{5}$$

wherein

$$
(6) \qquad
\begin{aligned}
a^{(21)} &= [L_v + \alpha + \beta N + L(N_u + \gamma L + \delta)]/(1 - LN),\\
a^{(12)} &= [N(L_v + \alpha + \beta N) + N_u + \gamma L + \delta]/(1 - LN),\\
b^{(21)} &= [\alpha_v + \beta\gamma + L(\alpha\gamma + \gamma_u + q)]/(1 - LN),\\
c^{(12)} &= [N(\beta\delta + \beta_v + p) + \beta\gamma + \delta_u]/(1 - LN).
\end{aligned}
$$

The points ρ, σ, τ are covariant points with respect to all transformations of proportionality factor and independent variable in equations (1) that preserve the net.

The curvilinear differential equation defining the asymptotic net on the surface is

$$
(7) \qquad Ldu^2 + 2dudv + Ndv^2 = 0.
$$

Evidently $L = N = 0$ is a necessary and sufficient condition that the parametric net be asymptotic. We also see that the surface is developable if, and only if, $1 - LN = 0$.

In order that the net defined by the equation

$$
(dv - \lambda du)(dv - \mu du) = 0 \qquad (\lambda \neq \mu)
$$

shall be a conjugate net the two directions λ and μ must separate harmonically the asymptotic directions. A necessary and sufficient condition for this is

$$
(8) \qquad \mu = -(L + \lambda)/(N\lambda + 1).
$$

The conjugate net whose directions separate harmonically the parametric directions is called by Green the *associate conjugate net*, and it is defined by the following equation:

$$
(9) \qquad Ldu^2 - Ndv^2 = 0.
$$

We list for reference the following invariants due to Green and Grove:*

$$
(10) \qquad
\begin{aligned}
r &= L + \lambda, \quad s = N\lambda + 1,\\
t &= L\lambda' - L'\lambda, \quad \pi = c_v - b_{1u},\\
h' &= q + b_1(\delta - b_1) - b_{1v} + Nb_{1u},\\
k &= p + c(\alpha - c) - c_u + Lc_v,\\
\mathfrak{B} &= Nc - \tfrac{1}{2}(a^{(12)} - \delta),\\
\mathfrak{C}' &= Lb_1 - \tfrac{1}{2}(a^{(21)} - \alpha),\\
G &= c^{(21)} + ca^{(21)} - c_u - c^2,\\
G' &= b^{(12)} + b_1 a^{(12)} - b_{1v} - b_1^2,\\
F &= c^{(12)} + ca^{(12)} - c\delta - Nc^2 - 2c_v,\\
F' &= b^{(21)} + b_1 a^{(21)} - b_1\alpha - Lb_1^2 - 2b_{1u},\\
W^{(u)} &= F - NG,
\end{aligned}
$$

* V. G. Grove, *A theory of a general net on a surface*, these Transactions, vol. 28 (1926), p. 496.

wherein $a^{(21)}$, $a^{(12)}$, $b^{(21)}$, $c^{(12)}$ are given by (6) and where

$$\begin{aligned} a &= 1/L, \quad b = -\alpha/L, \quad c = -\beta/L, \quad d = -p/L, \\ a_1 &= 1/N, \quad b_1 = -\gamma/N, \quad c_1 = -\delta/N, \quad d_1 = -q/N, \\ b^{(12)} &= [N(\alpha_v + \beta\gamma) + \alpha\gamma + \gamma_u + q]/(1 - LN), \\ c^{(21)} &= [\beta\delta + \beta_v + p + L(\beta\gamma + \delta_u)]/(1 - LN). \end{aligned}$$

3. **The non-conjugate osculating quadrics.** We now compute the equations of the quadrics defined in §1. Let us consider a non-conjugate net N on a surface S of the kind specified in §2, and suppose S is an integral surface of the system (1). On this surface let us consider a curve C_λ which is a member of the family defined by the equation

$$dv - \lambda du = 0, \tag{11}$$

and suppose this family of curves is not conjugate to either family of the net N. We shall first find the equation of the osculating quadric Q_u determined by the tangents to three consecutive u-curves at a point P_x of C_λ. The equation will first be referred to the local tetrahedron of reference whose vertices are the points x, x_u, x_v, x_{uv}. We may define any point X on C_λ near P_x by the following power series in Δu:

$$X = x + x'\Delta u + \frac{x''\Delta u^2}{2} \cdots \qquad \left(x' = \frac{dx}{du}\right).$$

By making use of the differential equations (1) and expressions obtained from them by differentiation this series can be expressed in the form

$$X = x_1 x + x_2 x_u + x_3 x_v + x_4 x_{uv},$$

where, for a suitably chosen unit point, x_1, x_2, x_3, x_4 are local coordinates of the point X referred to the local tetrahedron x, x_u, x_v, x_{uv}, and the power series defining them sufficiently far for the purposes of this paper are

$$\begin{aligned} x_1 &= 1 + \cdots, \\ x_2 &= \Delta u + \cdots, \\ x_3 &= \lambda\Delta u + (\beta + \lambda' + \lambda^2\delta)\Delta u^2/2 + \cdots, \\ x_4 &= (L + 2\lambda + N\lambda^2)\Delta u^2/2 + \cdots. \end{aligned} \tag{12}$$

Similarly for a point X_u on the u-tangent near P_x we get the local power series expansions

$$
(13)\qquad
\begin{aligned}
y_1 &= p\Delta u + \cdots,\\
y_2 &= 1 + \alpha\Delta u + \cdots,\\
y_3 &= \beta\Delta u + [\beta_u + \alpha\beta + (L + 2\lambda)c^{(21)}\\
&\quad + \lambda^2 c^{(12)}]\Delta u^2/2 + \cdots,\\
y_4 &= (L + \lambda)\Delta u + [\lambda' + L_u + \beta + \alpha L\\
&\quad + (L + 2\lambda)a^{(21)} + \lambda^2 a^{(12)}]\Delta u^2/2 + \cdots.
\end{aligned}
$$

Now let us set

$$
\begin{aligned}
e_1 &= (\beta + \lambda' + \lambda^2\delta)/2,\\
e_2 &= (L + 2\lambda + N\lambda^2)/2,\\
e_3 &= [\beta_u + \alpha\beta + (L + 2\lambda)c^{(21)} + \lambda^2 c^{(12)}]/2,\\
e_4 &= [\lambda' + L_u + \beta + \alpha L + (L + 2\lambda)a^{(21)} + \lambda^2 a^{(12)}]/2.
\end{aligned}
$$

Any point on the line XX_u is given by the linear combination

$$
mX + nX_u \qquad (m, n \text{ scalars}).
$$

On making use of equations (12) and (13) the power series expansions for the local coordinates of the point are found to be

$$
(14)\qquad
\begin{aligned}
z_1 &= m + pn\Delta u + \cdots,\\
z_2 &= n + (m + \alpha n)\Delta u + \cdots,\\
z_3 &= (\lambda m + \beta n)\Delta u + (e_1 m + e_3 n)\Delta u^2 + \cdots,\\
z_4 &= rn\Delta u + (e_2 m + e_4 n)\Delta u^2 + \cdots.
\end{aligned}
$$

Now demand that the power series (14) satisfy the most general equation of a quadric surface identically in m, n and identically in Δu as far as terms of the second degree. The computation involved in doing this is rather laborious, and the details will be omitted here. The equation of the quadric Q_u is found to be

$$
(15)\qquad (L - N\lambda^2)x_3^2 + Rx_3x_4 + S'x_4^2 + 2\lambda(\lambda x_1x_4 - rx_2x_3 + \beta x_2x_4) = 0,
$$

where r was given in equations (10) and R and S' are given by

$$
\begin{aligned}
R &= 2\lambda(\mathfrak{C}' - Lb_1) + 2\lambda^2(\mathfrak{B} - Nc) + [\lambda' - 2\beta - \lambda(\log r)'],\\
S' &= (\beta^2 + \lambda\beta' - \alpha\beta\lambda - p\lambda^2 - \beta\lambda')/r + \lambda[c^{(21)} + \lambda c^{(12)}].
\end{aligned}
$$

The equation of the osculating quadric Q_v may be obtained immediately from that of Q_u by performing in equation (15) the following substitution:

$$\begin{pmatrix} u & x_2 & \alpha & \beta & p & L & \lambda & \lambda' \\ v & x_3 & \delta & \gamma & q & N & 1/\lambda & -\lambda'/\lambda^3 \end{pmatrix}. \tag{16}$$

The resulting equation of Q_v is

$$(L - N\lambda^2)x_2^2 + Px_2x_4 + Qx_4^2 - 2(x_1x_4 - sx_2x_3 + \gamma\lambda x_3x_4) = 0, \tag{17}$$

where s was given in equations (10) and P and Q are defined by

$$P = 2\gamma\lambda^2 + (\log s)' - 2(\mathfrak{C}' - Lb_1) - 2\lambda(\mathfrak{B} - Nc),$$
$$Q = \lambda[\delta\gamma\lambda - \gamma' + q - \gamma(\log\lambda)' - \gamma^2\lambda^2]/s - (b^{(21)} + \lambda b^{(12)}).$$

If we specialize our net to be the asymptotic net by putting $L = N = 0$, equations (15) and (17) reduce to the equations for the asymptotic osculating quadrics of Bompiani and Klobouček.

4. **A covariant tetrahedron of reference.** The tetrahedron to which the equations of the osculating quadrics were referred in the last section was not a covariant tetrahedron, and hence we were unable to express the coefficients in the equations completely in terms of invariants. We find it advantageous to change our tetrahedron of reference to a new covariant tetrahedron whose vertices are the points x, ρ, σ, τ, of which the last three were defined by equations (2), (3), (4) of §2. The transformation that effects this change of reference system is

$$x_1 = y_1 + (\beta/L)y_2 + (\gamma/N)y_3 + K'y_4,$$
$$x_2 = y_2 + (\gamma/N)y_4,$$
$$x_3 = y_3 + (\beta/L)y_4,$$
$$x_4 = y_4,$$

where the x's are the old variables and the y's are the new. After renaming the variables so that the x's are the new local coordinates, the transformed equations of Q_u and Q_v are respectively

$$\begin{aligned} (L - N\lambda^2)x_3^2 &+ (2\lambda^2\mathfrak{B} + 2\lambda\mathfrak{C}' + t/r)x_3x_4 + \lambda(G - \lambda F' \\ &+ \lambda\pi - \lambda k/r)x_4^2 + 2\lambda^2x_1x_4 - 2\lambda r x_2x_3 = 0, \\ (L - N\lambda^2)x_2^2 &- [2\lambda\mathfrak{B} + 2\mathfrak{C}' - (\log s)']x_2x_4 + (F - \lambda G' \\ &+ \lambda\pi + \lambda h'/s)x_4^2 - 2x_1x_4 + 2sx_2x_3 = 0. \end{aligned} \tag{18}$$

The coefficients in the equations of the osculating quadrics are now expressed completely in terms of invariants defined in (10).

5. **Some immediate results.** In this section we establish some theorems

analogous to those of Lane for a conjugate net and in addition obtain an interesting reciprocity property.

The osculating quadric Q_u intersects the tangent plane, $x_4=0$, in the u-tangent, $x_3=x_4=0$, and the residual line

$$x_4 = (L - N\lambda^2)x_3 - 2\lambda r x_2 = 0. \tag{19}$$

This line is the v-tangent if, and only if, $L-N\lambda^2=0$. Likewise Q_v intersects the tangent plane in the v-tangent, $x_2=x_4=0$, and the residual line

$$x_4 = (L - N\lambda^2)x_2 + 2sx_3 = 0. \tag{20}$$

On referring to equation (9) we conclude the following result:

A necessary and sufficient condition that a curve on our surface belong to the associate conjugate net is that either of the two osculating quadrics have the tangents of the net as generators at every point of the curve.

Now suppose $L-N\lambda^2\neq 0$. Then the two lines (19) and (20) coincide at every point of C_λ if, and only if, equation (7) holds. Thus we have the following theorem:

The residual intersections of the osculating quadrics with the tangent plane coincide at every point of a curve if, and only if, the curve is a member of one family of the asymptotic net on the surface. The line of coincidence is the tangent to the curve.

The coordinates ξ of the polar plane of any point x with respect to the osculating quadric Q_u are given by

$$\begin{aligned} \xi_1 &= 2\lambda^2 x_4, \quad \xi_2 = -\,2\lambda r x_3, \\ \xi_3 &= -\,2\lambda r x_2 + 2(L - N\lambda^2)x_3 \\ &\quad + (2\lambda^2\mathfrak{B} + 2\lambda\mathfrak{C}' + t/r)x_4, \\ \xi_4 &= 2\lambda^2 x_1 + (2\lambda^2\mathfrak{B} + 2\lambda\mathfrak{C}' + t/r)x_3 \\ &\quad + 2\lambda(G - \lambda F' + \lambda\pi - \lambda k/r)x_4. \end{aligned} \tag{21}$$

The polar plane of the point ρ, whose local coordinates are (0, 1, 0, 0), with respect to Q_u is the plane $x_3=0$, which is the osculating plane to the u-curve C_u at P_x. But we see from the first of equations (18) that ρ lies on Q_u, hence we conclude that *the osculating plane to the curve C_u at a point of a curve C_λ is tangent to the corresponding osculating quadric Q_u at the point ρ.* In a similar manner it may be shown that *the osculating plane to C_v at P_x is tangent to Q_v at the point σ.*

By referring to (21) we see that the polar plane of the point σ with respect to Q_u passes through the point P_x, and it passes also through the point τ if, and only if,

$$2\lambda^2\mathfrak{B} + 2\lambda\mathfrak{C}' + t/r = 0. \tag{22}$$

We readily draw the following conclusion:

Condition (22) *is necessary and sufficient for the ray and axis to be reciprocal polar lines with respect to the quadric* Q_u.

Similar results hold for the quadric Q_v.

6. Nature of the intersection of the osculating quadrics. We now make a complete study of the nature of the curve of intersection of the osculating quadrics Q_u and Q_v. This study is made by means of the elementary divisors of the matrix of the pencil of quadrics based on Q_u and Q_v. The method is to compute the elementary divisors, write down the characteristic, and ascertain the general nature of the intersection by referring to results tabulated by Snyder and Sisam.*

For the sake of brevity let us write equations (18) for the osculating quadrics Q_u and Q_v in the following form:

$$\begin{aligned} &Ax_3^2 + 2Bx_3x_4 + Cx_4^2 + 2\lambda^2x_1x_4 - 2\lambda rx_2x_3 = 0, \\ &Ax_2^2 - 2Dx_2x_4 + Ex_4^2 - 2x_1x_4 + 2sx_2x_3 = 0, \end{aligned} \tag{23}$$

where A, B, C, D, E are new notations introduced here, and their values may be readily read off from (18). We now give the results obtained for the various cases that arise in our problem.

Case I. If $A\neq 0$, $B-D\lambda\neq 0$, $L+2\lambda+N\lambda^2\neq 0$, we find that the characteristic is [13]. Hence the intersection of the quadrics Q_u and Q_v at each point of a curve C_λ satisfying the conditions stated is a quartic curve with a cusp. But if $B-D\lambda=0$ we get by using the values of B and D

$$t/r + \lambda(\log s)' = 0, \tag{24}$$

and

$$t/r = (L\lambda' - L'\lambda)/(L+\lambda) = \lambda' - \lambda[\log(L+\lambda)]'.$$

Substituting this in (24), dividing by λ, and integrating we get

$$\lambda(N\lambda + 1)/(L+\lambda) = e \qquad (e = \text{const.}).$$

Now by replacing λ by dv/du we see that if C_λ is a curve such that $B-D\lambda=0$ then C_λ belongs to the pencil of families of curves defined by the curvilinear differential equation

$$eLdu^2 + (e+1)dudv + Ndv^2 = 0. \tag{25}$$

But let us note that for $e=-1$ this becomes $L-N\lambda^2=0$, and for $e=+1$ it

* V. Snyder and C. H. Sisam, *Analytic Geometry of Space*, New York, 1914, p. 163.

becomes $L+2\lambda+N\lambda^2=0$ after dividing through by du^2. Thus the three conditions imposed at the beginning of this case are all implied by the second. Since in later work no other case arises in which the curve of intersection of the two quadrics is non-degenerate, we state the following theorem:

The two osculating quadrics at each point of a curve on our net intersect in a non-composite quartic curve if, and only if, the curve does not belong to the pencil of families of curves defined by equation (25). *The quartic has a cusp at the point* P_x.

In our general treatment *no case arises in which at each point of a curve on our net the total intersection of the pair of osculating quadrics contains as a part of it a non-degenerate cubic.*

Case II (a). Suppose $B-D\lambda=0$, $A\neq 0$, $L+2\lambda+N\lambda^2\neq 0$, $A(C+E\lambda^2)-B^2\neq 0$. We get the characteristic $[1(21)]$, which shows that the curve of intersection is composed of two conics which touch each other. *The conics touch at* P_x *and lie in the planes whose equations are*

$$A\lambda(x_3-\lambda x_2)+[B\lambda\pm(B^2-A(C+E\lambda^2))^{1/2}]x_4=0. \tag{26}$$

Case II (b). If in II(a) we have $A(C+E\lambda^2)-B^2=0$, then the characteristic is $[1(111)]$, and hence the intersection is a conic counted twice. *The equations of this conic are found to be*

$$\begin{aligned} A(x_3-\lambda x_2)+Bx_4&=0,\\ A(L+2\lambda+N\lambda^2)x_3^2+2B\lambda s x_3x_4+ACx_4^2+2A\lambda^2x_1x_4&=0. \end{aligned} \tag{27}$$

Case II (c). If $L+2\lambda+N\lambda^2=0$, $A(C+E\lambda^2)-B^2\neq 0$, the characteristic is $[(22)]$, and the intersection is three generators, one counted twice. *The generator that is counted twice is the line* $x_4=x_3-\lambda x_2=0$, *which is tangent to the curve* C_λ *at* P_x. *The other two generators are the lines whose equations are as follows:*

$$\begin{aligned} A\lambda(x_3-\lambda x_2)+[B\lambda\pm(B^2-A(C+E\lambda^2))^{1/2}]x_4&=0,\\ A\lambda(Cx_4+2\lambda^2x_1)-[B\lambda^2s\mp(B^2-A(C+E\lambda^2))^{1/2}]x_3&=0, \end{aligned} \tag{28}$$

where the upper signs in the two equations are paired to give one generator and the lower signs the other.

Case II (d). If now in II(c) we let $A(C+E\lambda^2)-B^2=0$, then the characteristic is $[(211)]$, and the intersection is two intersecting generators, each counted twice. *The equations of these generators are respectively*

$$\begin{aligned} x_4=x_3-\lambda x_2&=0,\\ A(x_3-\lambda x_2)+Bx_4=2A\lambda^2x_1+2B\lambda s x_3+ACx_4&=0. \end{aligned} \tag{29}$$

From the above results we note the following theorem:

At each point of an asymptotic curve on the net the intersection of the osculating quadrics degenerates completely into generators.

Case III. Now let $A=0$, $B\neq 0$. The characteristic is then $[(31)]$, and the intersection is two intersecting generators, *the tangents of our net at P_x*, and a residual conic. We thus obtain the following result by using (22):

If the ray and axis are not reciprocal polar lines with respect to the osculating quadric Q_u, and if the curve C_λ belongs to the associate conjugate net $(A=0)$, then the quadrics Q_u and Q_v intersect in the tangents of the net and a residual non-degenerate conic. The equations of this conic are

$$(30)\qquad \begin{aligned} 2D\lambda^2 x_2 - 2Bx_3 - (C + E\lambda^2)x_4 &= 0,\\ CDx_4^2 + [2BD - s(C + E\lambda^2)]x_3x_4 - 2Bsx_3^2 + 2D\lambda^2 x_1x_4 &= 0. \end{aligned}$$

Case IV. If $A=B=0$, $C+E\lambda^2\neq 0$, we find that the characteristic is $[(211)]$, and *the intersection is composed of the tangents of the net counted twice. The quadrics touch at each point of each tangent.*

Case V. If $A=B=C+E\lambda^2=0$, the characteristic is $[(1111)]$, and *the two quadrics Q_u and Q_v coincide. Their equations both become*

$$(31)\qquad Cx_4^2 + 2\lambda^2 x_1x_4 - 2\lambda r x_2x_3 = 0.$$

This completes the discussion of all cases that arise.

7. **Conjugate nets.** Let us now consider two curves C_λ and C_μ on our net, and let us suppose that they are respectively embedded in two conjugate one-parameter families of curves on the net. By equation (8) we see that a necessary and sufficient condition for this is $\mu=-r/s$. Let P_x denote the point where C_λ crosses C_μ. We shall denote by $Q_{u\lambda}$ the osculating quadric Q_u of C_λ at P_x and by $Q_{u\mu}$ the osculating quadric Q_u of C_μ at P_x. The equation of $Q_{u\mu}$ may be obtained from that of $Q_{u\lambda}$ simply by replacing λ by $-r/s$. The equations of the osculating quadrics $Q_{u\lambda}$ and $Q_{u\mu}$ are thus as follows:

$$(32)\qquad \begin{aligned} Ax_3^2 + 2Bx_3x_4 + Cx_4^2 + 2\lambda^2 x_1x_4 - 2\lambda r x_2x_3 &= 0,\\ A(1-LN)x_3^2 + 2Hx_3x_4 + Kx_4^2 + 2r^2x_1x_4 - 2\lambda r(1-LN)x_2x_3 &= 0, \end{aligned}$$

where the first of these is the same as the first of equations (18) and H and K are defined by

$$\begin{aligned} H = r^2\mathfrak{B} - rs\mathfrak{C}' + s^2[L(rs' - r's)&\\ + r(L_u s - L_v r)]/[2\lambda s(LN-1)],&\\ K = r\{\pi r - rF' - sG - (rsk)/[\lambda(LN-1)]\}.& \end{aligned}$$

The cone projecting the curve of intersection of the two quadrics (32) *from the second vertex of the tetrahedron of reference is given by*

$$(33)\quad \begin{aligned} 2[B(1 - LN) - H]x_3x_4 + [C(1 - LN) - K]x_4^2 \\ + 2[\lambda^2(1 - LN) - r^2]x_1x_4 = 0. \end{aligned}$$

This cone evidently consists of two planes, one of which is the tangent plane, $x_4=0$. But the tangent plane intersects each of the quadrics (32) in the two lines

$$(34)\quad \begin{aligned} x_3 = x_4 = 0, \\ x_4 = Ax_3 - 2\lambda r x_2 = 0. \end{aligned}$$

Hence we draw the following conclusion:

At any point on the net the osculating quadric $Q_{u\lambda}$ of a curve of one family of a conjugate net always intersects the corresponding quadric $Q_{u\mu}$ of the curve of the other family in two conics, one of which is degenerate and is represented by equations (34). *The other conic lies in the plane*

$$(35)\ 2L(L + 2\lambda + N\lambda^2)x_1 - 2[B(1 - LN) - H]x_3 - [C(1 - LN) - K]x_4 = 0.$$

Let us note that since C_λ and C_μ are assumed distinct it follows as a consequence that $L+2\lambda+N\lambda^2 \neq 0$, and previous restrictions insure that $L\lambda r \neq 0$. Now let us project the curve of intersection of the quadrics (32) from the first vertex, P_x, of the tetrahedron of reference. *The equation of the projecting cone is*

$$(36)\quad \begin{aligned} A[r^2 - \lambda^2(1 - LN)]x_3^2 + 2(r^2B - \lambda^2H)x_3x_4 + \{[L\lambda r(L + 2\lambda \\ + N\lambda^2)W^{(u)}]/(1 - LN)\}x_4^2 - 2\lambda r[r^2 - \lambda^2(1 - LN)]x_2x_3 = 0. \end{aligned}$$

Evidently the lines (34) are generators of this cone; so this is really the cone projecting from P_x the conic lying in the plane (35), which does not pass through P_x. Hence the condition for the cone (36) to degenerate is just the condition for the residual conic to degenerate. But evidently $W^{(u)}=0$ is necessary and sufficient for this, and, as Green has shown, this is just the condition that insures that the u-tangents of the net form a W-congruence; that is, that the asymptotic curves on the two focal sheets of the congruence correspond. Hence we have proved the following theorem:

A necessary and sufficient condition that the congruence of tangents to the curves C_u of the net be generators of a W-congruence is that the residual conic of intersection of the two quadrics $Q_{u\lambda}$ and $Q_{u\mu}$ degenerate at every point of the net.

We have already seen that if the curve C_μ is embedded in a one-parameter family of curves that is conjugate to the family to which C_λ belongs, then the

osculating quadrics $Q_{u\lambda}$ and $Q_{u\mu}$ each intersect the tangent plane, $x_4=0$, in the lines (34); that is, in a u-tangent and a residual line. Now let us suppose that C_λ and C_μ are any two curves which are respectively members of two distinct one-parameter families of curves, neither of which coincides with either family of the net itself. Then the corresponding osculating quadrics Q_u for these two curves ordinarily will intersect the tangent plane in the u-tangent, $x_3=x_4=0$, and the two residual lines

$$(37)\qquad \begin{aligned} x_4 &= (L - N\lambda^2)x_3 - 2\lambda(L+\lambda)x_2 = 0,\\ x_4 &= (L - N\mu^2)x_3 - 2\mu(L+\mu)x_2 = 0. \end{aligned}$$

Now if we demand that these two residual lines coincide we get $\mu=-r/s$. Hence we draw the following conclusion:

A necessary and sufficient condition that a net on our surface be a conjugate net is that the residual intersections (37) *of the quadrics $Q_{u\lambda}$ and $Q_{u\mu}$ with the tangent plane coincide at every point of the net.*

8. **Metric considerations.** It is the purpose of this section to indicate a method of studying the analogue of the preceding problem in metric three-space and to give a few results that have been obtained by the author for the metric situation. Throughout this section we shall employ the notation and vector methods of Blaschke* for studying the geometry of a surface in ordinary space.

The equations of the osculating quadrics at the point x of a curve on the surface are here referred to a local orthogonal cartesian coordinate system, which is defined by the convention that any point X given by an expression of the form

$$(38)\qquad X - x = y_1x_u/E^{1/2} + y_2x_v/G^{1/2} + y_3\xi,$$

where ξ is the unit normal at the point x, shall have local coordinates y_1, y_2, y_3.

If we now let the partial differential equations of Gauss and Weingarten play the role of equations (1) and compute the equation of Q_u in a manner similar to that of §3, we find that *the equation of Q_u is*

$$(39)\qquad \begin{aligned} E^{1/2}(L - N\lambda^2)y_2^2 + SGE^{1/2}y_3^2 - 2QG^{1/2}y_1y_2 &\\ + 2PGy_1y_3 + 2S(EG)^{1/2}y_2y_3 + 2\lambda GE^{1/2}y_3 &= 0, \end{aligned}$$

where

* W. Blaschke, *Vorlesungen über Differentialgeometrie*, 3d edition, vol. 1, Berlin, 1930, pp. 85–118.

$$P = \begin{Bmatrix} 11 \\ 2 \end{Bmatrix} + \lambda \begin{Bmatrix} 12 \\ 2 \end{Bmatrix}, \quad Q = L + M\lambda,$$

$$\begin{aligned} \lambda QS = {} & \begin{Bmatrix} 11 \\ 2 \end{Bmatrix}^2 - \lambda' \begin{Bmatrix} 11 \\ 2 \end{Bmatrix} + \lambda \left(2 \begin{Bmatrix} 11 \\ 2 \end{Bmatrix} \begin{Bmatrix} 12 \\ 2 \end{Bmatrix} - \begin{Bmatrix} 11 \\ 1 \end{Bmatrix} \begin{Bmatrix} 11 \\ 2 \end{Bmatrix} + \begin{Bmatrix} 11 \\ 2 \end{Bmatrix}_u \right. \\ & \left. + L(FL - EM)/W^2 \right) + \lambda^2 \left(\begin{Bmatrix} 11 \\ 2 \end{Bmatrix} \begin{Bmatrix} 22 \\ 2 \end{Bmatrix} - 2 \begin{Bmatrix} 11 \\ 2 \end{Bmatrix} \begin{Bmatrix} 12 \\ 1 \end{Bmatrix} + 2 \begin{Bmatrix} 11 \\ 2 \end{Bmatrix}_v \right. \\ & \left. + 2L(FM - EN)/W^2 \right) + \lambda^3 \left(\begin{Bmatrix} 11 \\ 2 \end{Bmatrix} \begin{Bmatrix} 22 \\ 1 \end{Bmatrix} - 2 \begin{Bmatrix} 12 \\ 1 \end{Bmatrix} \begin{Bmatrix} 12 \\ 2 \end{Bmatrix} + \begin{Bmatrix} 22 \\ 2 \end{Bmatrix}_u \right. \\ & \left. + N(FL - EM)/W^2 \right), \end{aligned}$$

$$\begin{aligned} 2\lambda QT = {} & L\lambda' - 2L \begin{Bmatrix} 11 \\ 2 \end{Bmatrix} + \lambda \left(L \begin{Bmatrix} 11 \\ 1 \end{Bmatrix} - L \begin{Bmatrix} 12 \\ 2 \end{Bmatrix} - 2M \begin{Bmatrix} 11 \\ 2 \end{Bmatrix} - L_u \right) \\ & + \lambda^2 \left(L \begin{Bmatrix} 22 \\ 2 \end{Bmatrix} + 2L \begin{Bmatrix} 12 \\ 1 \end{Bmatrix} - N \begin{Bmatrix} 11 \\ 2 \end{Bmatrix} - 2L_v \right) \\ & + \lambda^3 \left(N \begin{Bmatrix} 12 \\ 2 \end{Bmatrix} + 2M \begin{Bmatrix} 12 \\ 1 \end{Bmatrix} - L \begin{Bmatrix} 22 \\ 1 \end{Bmatrix} - N_u \right). \end{aligned}$$

The equation of the osculating quadric Q_v may now be written down by symmetry, or it may be obtained by direct calculation.

The discriminant Δ of the quadric (39) is

$$\Delta = \lambda^2 E G^3 Q^2 \neq 0,$$

for E, G are positive on a real surface, $\lambda \neq 0$ since C_λ was assumed not a u-curve, and $Q \neq 0$ since C_λ was supposed not a curve of the family conjugate to the family of u-curves of the net. *Thus the osculating quadric Q_u is always non-singular for a real analytic surface.*

If we simplify the analysis by taking the parametric net to be orthogonal we get the following interesting result. The determinant D' of the matrix of the second degree terms in the resulting equation of the quadric Q_u then becomes

$$D' = (E^{1/2}/\lambda)(A_0\lambda' + A_1 + A_2\lambda + A_3\lambda^2 + A_4\lambda^3), \tag{40}$$

wherein

$$A_0 = -G(LG_u + ME_v)/2,$$

$$A_1 = (4L^2MG + 2LGE_{uv} + ME_v^2 - LE_vG_u - 2L_uE_vG)/4,$$

$$A_2 = (2GL_uG_u + 8L^2NG + 4LGE_{vv} - 2LE_vG_v - MGE_uE_v/E + 4LM^2G$$
$$+ 2NE_v^2 + 2MGE_{uv} - 2E_vG_u - 4GL_vE_v - LGE_uG_u/E)/4,$$
$$A_3 = (12LMNG - 2LGG_{uv} + 4MGE_{vv} - 3ME_vG_v - 2GN_uE_v + LG_uG_v$$
$$- 2NE_vG_u)/4,$$
$$A_4 = (4M^2NG - MGE_vG_u/E - 2MGG_{uv} + 2MG_uG_v - LGG_u^2/E + 2GN_uG_u)/4.$$

Now suppose C_λ is a curve such that $A_0 \neq 0$. Since by definition the rulings of the quadric Q_u are real, we find that *the osculating quadric Q_u at every point of a curve C_λ on an orthogonal net is a hyperbolic paraboloid if, and only if, C_λ belongs to the family of hypergeodesics defined by $D' = 0$. Otherwise Q_u is a hyperboloid.*

LAMBUTH COLLEGE,
JACKSON, TENN.

ON RIESZ AND CESÀRO METHODS OF SUMMABILITY*

BY

RALPH PALMER AGNEW†

1. Introduction. Marcel Riesz‡ formulated the following method of summability: Let r be any complex constant and, given a series $u_0+u_1+u_2+\cdots$, let

$$(1.1)\qquad A_r:\quad \alpha_n = \sum_{k=0}^{n-1}\left(1-\frac{k}{n}\right)^r u_k \qquad (n = 1, 2, 3, \cdots);$$

if $\lim_{n\to\infty}\alpha_n = L$, then $\sum u_n$ is said to be summable A_r to L.

In a second note, Riesz§ gave a method which is, when $r>0$, equivalent‖ (vide Theorem 4.4) to the following: Let r be a complex constant and let¶

$$(1.2)\qquad B_r:\quad \beta(t) = \sum_{k=0}^{[t]-1}\left(1-\frac{k}{t}\right)^r u_k, \qquad 1 \leqq t < \infty;$$

if $\beta(t)$ approaches a limit L as t becomes infinite over the real set $t\geqq 1$, then $\sum u_n$ is summable B_r to L. The second method of Riesz is the following: Let $\mathcal{R}(r)>0$, and let

$$(1.3)\qquad D_r:\quad \delta(t) = \sum_{k=0}^{[t^-]}\left(1-\frac{k}{t}\right)^r u_k, \qquad 1 \leqq t < \infty;$$

if $\delta(t)$ converges to L as t becomes infinite continuously, then $\sum u_n$ is summable D_r to L. The method D_r is known as the Riesz method of order r and type $\lambda_n = n$**, and has proved to be one of the most useful of all methods of summability.

In his second note, Riesz outlined a proof that D_r is equivalent to C_r,

---

* Presented to the Society, December 30, 1931; received by the editors August 24, 1932.

† National Research Fellow.

‡ Comptes Rendus, vol. 149 (1909), pp. 18–22. In this note Riesz considered only real positive orders r.

§ Comptes Rendus, vol. 152 (1911), pp. 1651–1654. Here again Riesz considered only the case $r>0$.

‖ The terminology used in this paper is that given by W. A. Hurwitz, Bulletin of the American Mathematical Society, vol. 28 (1922), pp. 17–36.

¶ We use the symbols $[t]$ and $[t^-]$ to denote respectively the greatest integer $\leqq t$ and the greatest integer $<t$.

** Hardy-Riesz, *General Theory of Dirichlet's series*, Cambridge Tracts in Mathematics and Mathematical Physics, No. 18.

the Cesàro method of order r, when $r>0$. Chapman* has stated that Riesz's proof of equivalence of D_r and C_r holds when $r>-1$; but this statement is incorrect as Theorem 2.1 shows. Hobson† has given a more detailed proof of equivalence of D_r and C_r when $r>0$.

In a third note, Riesz‡ outlined a proof that A_r and C_r are equivalent when $-1<r<1$, and showed that this equivalence does not hold for certain values of $r>1$.

It is the object of this paper to discuss A_r, B_r, D_r, C_r, and closely related methods of summability. We shall be especially interested in orders r with real part $\mathfrak{R}(r)<0$.

In §2 we show that D_r does not constitute a useful method of summability when $\mathfrak{R}(r)<0$; and in §§2–3 we discuss modifications of D_r which may be expected to be useful when $\mathfrak{R}(r)<0$. For each complex r, these modifications are found to be equivalent to B_r. In §4 we show that B_r and D_r are equivalent when $r\geqq 0$. In §§5–7 we obtain auxilliary results from which it follows in §8 that B_r and C_r are equivalent when $-1<\mathfrak{R}(r)<0$. The theorems of §8 give a complete solution of the problem which furnished the point of departure of this investigation. In §9 we give relations between methods B_r of different orders. We show in §§10–11 that A_r does not possess certain properties of B_r when $\mathfrak{R}(r)<-1$; in particular when $\mathfrak{R}(r)$ is less than a certain constant between -2 and -1, A_r is not consistent with convergence. Finally, in §12 we point out that when $\mathfrak{R}(r)<0$, the methods A_r, B_r, and C_r are equivalent over a certain class of series.

2. Ineffectiveness of D_r when $\mathfrak{R}(r)<0$. It is well known that D_r is regular when r is real and $\geqq 0$. It can be shown further that D_r is regular when $\mathfrak{R}(r)>0$; and that D_r is not regular when $\mathfrak{R}(r)=0$ but $r\neq 0$. To show that D_r does not constitute a useful method of summability when $\mathfrak{R}(r)<0$, we will prove the following Theorem.

THEOREM 2.1. *In order that $\sum u_n$ may be summable D_r when $\mathfrak{R}(r)<0$, it is necessary and sufficient that $\sum u_n$ have at most a finite number of terms different from zero.*

Sufficiency is easily established. To prove necessity, let us suppose that $u_p\neq 0$ for a certain index $p>0$; then $\lim_{h\to 0+}|\delta(p+h)|=+\infty$. Hence if $u_k\neq 0$ for an infinite set of values of k, then $\delta(t)$ is unbounded over every interval

* Proceedings of the London Mathematical Society, (2), vol. 9 (1910–11), p. 374, second footnote.

† *Theory of Functions of a Real Variable.* vol. II, 1926, pp. 90–98.

‡ Proceedings of the London Mathematical Society, (2), vol. 22 (1923–24), p. 418.

(N, ∞) so that $\delta(t)$ cannot converge as $t\to\infty$ and the theorem is proved.*

Theorem 2.1 and its proof make it clear that if a useful generalization to orders with real part $\Re(r)<0$ of the Riesz method D_r is to be obtained, the upper index of summation with respect to k must be a function of t which is definitely less than $[t^-]$. The two methods defined by the transformations

$$\pi(t) = \sum_{k=0}^{[t-\theta]}\left(1-\frac{k}{t}\right)^r u_k, \qquad \theta < t < \infty, \tag{2.2}$$

$$\rho(t) = \sum_{k=0}^{[(t-\theta)^-]}\left(1-\frac{k}{t}\right)^r u_k, \qquad \theta < t < \infty, \tag{2.3}$$

where θ is a positive constant, suggest themselves at once as modifications of D_r which may be useful for every complex order.

Let r be any complex number. Then, corresponding to any given series $\sum u_n$, the functions $\pi(t)$ and $\rho(t)$ are equal except when t is of the form $t=n+\theta$ and $u_n\neq 0$, in which case $\pi(n+\theta)\neq\rho(n+\theta)$. Furthermore the transforms $\pi(t)$ and $\rho(t)$ are continuous except when $t=n+\theta$ and $u_n\neq 0$; here $\pi(t)$ and $\rho(t)$ have finite jumps, $\pi(t)$ having right-hand continuity and $\rho(t)$ having left-hand continuity. It follows that if either $\pi(t)$ or $\rho(t)$ converges as $t\to\infty$, then the other must also converge to the same value as $t\to\infty$. Hence the methods (2.2) and (2.3) are equivalent. We elect to consider the first rather than the second of these.

3. Consideration of (2.2) for different values of θ. In this section we will establish a theorem which will be of fundamental importance in the sequel; and will show that, for any fixed complex r, the methods (2.2) obtained by selecting different positive values of θ are equivalent to B_r.

THEOREM 3.1. *If $\sum u_n$ is summable* (2.2) *with r a fixed complex constant and θ a fixed positive constant, then*

$$\lim_{n\to\infty} u_n/n^r = 0. \tag{3.11}$$

Suppose $\sum u_n$ is summable (2.2) to L. Then

$$\lim_{t\to\infty}\sum_{k=0}^{[t-\theta]}\left(1-\frac{k}{t}\right)^r v_k = 0$$

* There is an apparent inconsistency between Theorem 2.1 and Chapman's statement (loc. cit., p. 401) that the Dirichlet series $1^{-s}-2^{-s}+3^{-s}-4^{-s}+\cdots$, $s>0$, is summable $(R, n, -r)$, i.e. D_{-r}, when $r<s$. The last equation of p. 399 shows that Chapman has used the transformation B_r rather than D_r; and furthermore the second equation of p. 400 is correct only when $[n]=n$. Therefore, as a matter of fact, Chapman has not shown that $\sum(-1)^n(n+1)^{-s}$ is summable D_{-r} when $r<s$; what he has shown is that the series is summable A_{-r} when $r<s$. However, it follows from this result and Theorem 12.1 that the series $\sum(-1)^n(n+1)^{-s}$ is summable B_{-r} and C_{-r} as well as A_{-r} when $r<s$.

where $v_0 = u_0 - L$ and $v_n = u_n$ when $n > 0$. Given $\epsilon > 0$, choose $T > \theta$ such that

$$\left| \sum_{k=0}^{[t-\theta]} \left(1 - \frac{k}{t}\right)^r v_k \right| < \frac{\epsilon}{2}, \qquad t > T. \tag{3.12}$$

Let $n > T+1$, let $0 < h < 1$, and set $t = n + \theta - h$ in (3.12) to obtain

$$\left| \sum_{k=0}^{n-1} \left(1 - \frac{k}{n+\theta-h}\right)^r v_k \right| < \frac{\epsilon}{2}, \qquad 0 < h < 1. \tag{3.13}$$

The left member of (3.13) is a continuous function of h over the closed interval $0 \leqq h \leqq 1$; hence we may take the limit as $h \to 0$ to obtain

$$\left| \sum_{k=0}^{n-1} \left(1 - \frac{k}{n+\theta}\right)^r v_k \right| \leqq \frac{\epsilon}{2}, \qquad n > T + 1. \tag{3.14}$$

Again we may set $t = n + \theta$ in (3.12) and write the last term of the sum as a separate term to obtain

$$\left| \sum_{k=0}^{n-1} \left(1 - \frac{k}{n+\theta}\right)^r v_k + \left(\frac{\theta}{n+\theta}\right)^r v_n \right| < \frac{\epsilon}{2}, \qquad n > T + 1. \tag{3.15}$$

Combining (3.14) and (3.15), we find that $|\theta^r v_n/(n+\theta)^r| < \epsilon$ when $n > T+1$. Hence $\lim_{n\to\infty} \theta^r v_n/(n+\theta)^r = 0$ so that $\lim_{n\to\infty} v_n/n^r = 0$ and, since $v_n = u_n$ when $n > 0$, (3.11) follows. Thus Theorem 3.1 is proved.

A slight modification of the preceding argument shows that if $\sum u_n$ is bounded (2.2), then u_n/n^r is bounded for all $n > 0$.

THEOREM 3.2. *If r is a complex constant and*

$$\pi^{(\theta)}(t) = \sum_{k=0}^{[t-\theta]} \left(1 - \frac{k}{t}\right)^r u_k, \quad \pi^{(\theta')}(t) = \sum_{k=0}^{[t-\theta']} \left(1 - \frac{k}{t}\right)^r u_k$$

represent two different methods of the form (2.2) *with $\theta > 0$, $\theta' > 0$, and if furthermore* $\lim_{n\to\infty} u_n/n^r = 0$, *then*

$$\lim_{t\to\infty} \left\{ \pi^{(\theta)}(t) - \pi^{(\theta')}(t) \right\} = 0. \tag{3.21}$$

To establish this result, we may assume that $\theta > \theta'$ and show that the difference in the left member of (3.21) consists of a finite number of terms each of which approaches zero as $t \to \infty$.

From the two preceding theorems we obtain at once

THEOREM 3.3. *When r is any complex constant, the methods obtained by assigning different positive values to θ in* (2.2) *are equivalent.*

For if $\sum u_n$ is summable (2.2) for a positive value of θ, then $\lim u_n/n^r=0$ by Theorem 3.1; hence the hypotheses of Theorem 3.2 are satisfied and the conclusion (3.21) completes the proof of Theorem 3.3.

The only representative of the set of methods (2.2) which we will consider in the sequel is that for which $\theta=1$; in this case (2.2) becomes B_r.

4. Relations between B_r and D_r when $\Re(r)\geqq 0$. Before passing to a study of B_r when $\Re(r)<0$, we wish to point out that B_r is closely related to the familiar Riesz method D_r when $\Re(r)\geqq 0$.

THEOREM 4.1. *If $\Re(r)\geqq 0$ and $\lim u_n/n^r=0$, then*

$$\lim_{t\to\infty}\{\delta(t)-\beta(t)\}=0. \tag{4.11}$$

We find from (1.2) and (1.3) that $|\delta(t)-\beta(t)|\leqq|u_{[t]}/[t]^r|$ when $\Re(r)\geqq 0$ and $t>1$, and Theorem 4.1 follows.

THEOREM 4.2. *If $\Re(r)\geqq 0$, then D_r includes B_r.*

If $\sum u_n$ is summable B_r to L so that $\lim \beta(t)=L$, then $\lim u_n/n^r=0$ by Theorem 3.1 with $\theta=1$; hence the hypotheses of Theorem 4.1 are satisfied, the conclusion (4.11) shows that $\lim \delta(t)=L$, and Theorem 4.2 is proved.

THEOREM 4.3. *If $r\geqq 0$, then B_r includes D_r.*

The proposition being evident when $r=0$, we suppose $r>0$. Let $\sum u_n$ be summable D_r to L so that $\lim \delta(t)=L$. Then, using the fact (§1) that C_r includes D_r when $r>0$, we see that $\sum u_n$ must be summable C_r and hence, as is well known, that $\lim u_n/n^r=0$. Hence Theorem 4.1 shows that $\lim \beta(t)=L$ and Theorem 4.3 is proved.

Combining Theorems 4.2 and 4.3 with the fact that D_r and C_r are equivalent when $r\geqq 0$, we obtain

THEOREM 4.4. *If $r\geqq 0$, then B_r, D_r, and C_r are equivalent.*

5. A relation between the A_r and B_r transforms when $\Re(r)<0$. We proceed to establish some preliminary propositions, interesting in themselves, which will enable us to obtain relations between B_r and C_r.

THEOREM 5.1. *When $\Re(r)<0$, the assumption that*

$$\lim_{t\to\infty}\beta(t)=\lim_{t\to\infty}\sum_{k=0}^{[t]-1}\left(1-\frac{k}{t}\right)^r u_k=L \tag{5.11}$$

is equivalent to the two assumptions

$$\lim_{n\to\infty} u_n/n^r=0 \tag{5.12}$$

and

$$\lim_{n\to\infty} \alpha_n = \lim_{n\to\infty} \sum_{k=0}^{n-1} \left(1 - \frac{k}{n}\right)^r u_k = L.^* \tag{5.13}$$

That (5.11) implies (5.12) follows from Theorem 3.1 with $\theta = 1$; and that (5.11) implies (5.13) follows from the fact that $\alpha_n = \beta(n)$. Hence our problem here is to show that (5.12) and (5.13) together imply (5.11).

A consideration of the sequence v_n defined by $v_0 = u_0 - L$ and $v_n = u_n$, $n > 0$, shows that it is sufficient to prove (5.12) and (5.13) imply (5.11) when $L = 0$. We suppose therefore that $\Re(r) < 0$, that (5.12) holds, and that (5.13) holds with $L = 0$; we will show that (5.11) holds with $L = 0$.

Given $\epsilon > 0$, choose an index $N > 0$ so great that

$$\left| \sum_{k=0}^{n-1} \left(1 - \frac{k}{n}\right)^r u_k \right| < \frac{\epsilon}{2}, \qquad n \geqq N, \tag{5.14}$$

and

$$| u_n/n^r | < \epsilon/(4B | r |), \qquad n \geqq N, \tag{5.15}$$

where $B = 2^{-r'}(1 + \sum_{p=1}^{\infty} p^{r'-1})$ and $r' = \Re(r)$. Next, choose an index $P > N$ so great that

$$\frac{|r|}{t^2} \left| \sum_{k=0}^{N-1} k\left(1 - \frac{k}{t}\right)^{r-1} u_k \right| < \frac{\epsilon}{4}, \qquad t > P. \tag{5.16}$$

Let $n > P$ and consider the function

$$\beta(t) = \sum_{k=0}^{n-1} \left(1 - \frac{k}{t}\right)^r u_k, \qquad n \leqq t < n + 1. \tag{5.17}$$

Using (5.14), we see that

$$| \beta(n) | < \epsilon/2. \tag{5.18}$$

Differentiating (5.17) we find

$$\beta'(t) = \frac{r}{t^2} \sum_{k=0}^{n-1} k\left(1 - \frac{k}{t}\right)^{r-1} u_k, \qquad n \leqq t < n + 1,$$

where the derivative for $t = n$ is a right-hand derivative. Hence

$$\beta'(t) \leqq \frac{|r|}{t^2} \left| \sum_{k=0}^{N-1} k\left(1 - \frac{k}{t}\right)^{r-1} u_k \right| + \frac{|r|}{t^2} \sum_{k=N}^{n-1} k^{1+r'}\left(1 - \frac{k}{t}\right)^{r'-1} | u_k/k^r |,$$

and using (5.16) and (5.15) we obtain

---

* Consideration of independence of (5.12) and (5.13) is relegated to §10 where we study A_r.

$$|\beta'(t)| < \epsilon/4 + \epsilon\Phi_n(t)/(4B), \tag{5.19}$$

where

$$\Phi_n(t) = \sum_{k=1}^{n-1} \frac{k^{r'}(t-k)^{r'-1}}{t^{r'}}, \qquad n \leqq t < n+1. \tag{5.20}$$

But since $r'<0$,

$$\sum_{k=1}^{[t/2]} \frac{k^{r'}(t-k)^{r'-1}}{t^{r'}} \leqq \frac{1}{t} \sum_{k=1}^{[t/2]} \left(1 - \frac{k}{t}\right)^{r'-1} \leqq \frac{1}{t} \sum_{k=1}^{[t/2]} \left(\frac{1}{2}\right)^{r'-1} \leqq 2^{-r'},$$

and

$$\sum_{k=[t/2]+1}^{n-1} \frac{k^{r'}(t-k)^{r'-1}}{t^{r'}} \leqq \left(\frac{[t/2]+1}{t}\right)^{r'} \sum_{k=[t/2]+1}^{n-1} (t-k)^{r'-1} \leqq 2^{-r'} \sum_{p=1}^{\infty} p^{r'-1},$$

so that

$$\Phi_n(t) \leqq 2^{-r'}\left(1 + \sum_{p=1}^{\infty} p^{r'-1}\right) = B, \qquad n \leqq t < n+1. \tag{5.21}$$

From (5.19) and (5.21) we obtain

$$|\beta'(t)| < \epsilon/2, \qquad n \leqq t < n+1. \tag{5.22}$$

Using (5.18), (5.22), and the formula

$$|\beta(t)| \leqq |\beta(n)| + \int_n^t |\beta'(t)|\, dt, \qquad n \leqq t < n+1,$$

we find that $|\beta(t)| < \epsilon$, $n \leqq t < n+1$.

We have shown that if $n > P$, then $|\beta(t)| < \epsilon$, $n \leqq t < n+1$. It follows that if $t > P+1$, then $|\beta(t)| < \epsilon$. Hence $\lim \beta(t) = 0$ and Theorem 5.1 is proved.

6. Lemmas involving C_r. The Cesàro method C_r of order r (r not a negative integer) is defined by the transformation

$$C_r: \quad \gamma_n = \sum_{k=0}^{n} a_{nk} u_k \qquad (n = 0, 1, 2, \cdots), \tag{6.01}$$

where

$$a_{nk} = \frac{\Gamma(n+1)\Gamma(n-k+1+r)}{\Gamma(n+1+r)\Gamma(n-k+1)}, \qquad 0 \leqq k \leqq n. \tag{6.02}$$

The following two lemmas will be used in the next section.

LEMMA 6.1. *Corresponding to each complex constant* r *(not a negative integer) there is a bounded sequence* C_{nk} *of constants such that for each positive index* n *and each index* $k<n$

$$a_{nk} = \left(1 - \frac{k}{n}\right)^r \left(1 + \frac{C_{nk}}{n-k}\right). \tag{6.11}$$

Using the familiar asymptotic expansion of the logarithm of the gamma function of a complex argument,* we find

$$\log \{\Gamma(n+1+r)/\Gamma(n+1)\} = r \log n + H_n/n, \qquad n > 0, \tag{6.12}$$

where H_n is a bounded sequence of constants. Subtracting (6.12) from the relation obtained by replacing n by $n-k$ in it, we obtain

$$\log a_{nk} = r \log \{(n-k)/n\} - H_n/n + H_{n-k}/(n-k) \tag{6.13}$$

when $n>0$ and $k<n$. The lemma results from (6.13). The following lemma is easily deduced from (6.12).

LEMMA 6.2. *When* r *is not a negative integer*

$$\lim_{n\to\infty} n^r a_{nn} = \Gamma(1+r). \tag{6.21}$$

In §8 we shall need

LEMMA 6.3. *When* $\Re(r) \leqq -1$, r *not a negative integer, the condition* lim $u_n/n^r = 0$ *is not necessary in order that* $\sum u_n$ *may be summable* C_r.

The inverse of C_r is, when r is not an integer, given by

$$u_n = \sum_{k=0}^{n} (-1)^{n-k} \binom{k+r}{r} \binom{r+1}{n-k} \gamma_k \tag{6.31}$$

or

$$u_n = \sum_{k=0}^{n} \frac{\sin \pi r}{\pi} \frac{\Gamma(2+r)}{\Gamma(k+1)} \frac{\Gamma(k+1+r)}{\Gamma(1+r)} \frac{\Gamma(n-k-1-r)}{\Gamma(n-k+1)} \gamma_k. \tag{6.32}$$

Corresponding to each complex r which is not an integer, let the sequence $\gamma_n^{(r)}$ be defined by $\gamma_0^{(r)} = \pi/\{\sin \pi r \Gamma(2+r)\}$ and $\gamma_n^{(r)} = 0$ when $n>0$; and let $\sum u_n^{(r)}$ be the series whose C_r transform is $\gamma_n^{(r)}$. Substituting in (6.32) we find

$$u_n^{(r)} = \Gamma(n-1-r)/\Gamma(n+1) = n^{-2-r}(1+o(1)) \tag{6.33}$$

so that

---

* See, for example, J. L. W. V. Jensen (translation by T. H. Gronwall), Annals of Mathematics, (2), vol. 17 (1916), p. 136.

$$u_n^{(r)}/n^r = n^{-2(1+r)}(1 + o(1)). \tag{6.34}$$

The series $\sum u_n^{(r)}$ is summable C_r to 0 and, when $\Re(r) \leqq -1$, the right member of (6.34) fails to converge to 0 as $n \to \infty$; thus Lemma 6.3 is established.

7. A relation between the A_r and C_r transforms when $\Re(r) < 0$. With the lemmas of §6 at our disposal, we are in a position to prove the following theorem.

THEOREM 7.1. *If $\Re(r) < 0$ (r not a negative integer) and the terms of $\sum u_n$ satisfy the condition*

$$\lim_{n \to \infty} u_n/n^r = 0, \tag{7.11}$$

then

$$\lim_{n \to \infty} (\gamma_n - \alpha_n) = 0, \tag{7.12}$$

*where γ_n and α_n represent respectively the C_r and A_r transform of $\sum u_n$.**

Letting $\sum u_n$ be any series for which (7.11) holds, we have for each $n > 1$

$$\gamma_n - \alpha_n = a_{nn}u_n + \sum_{k=0}^{n-1} \left\{ a_{nk} - \left(1 - \frac{k}{n}\right)^r \right\} u_k. \tag{7.13}$$

Writing $a_{nn}u_n$ in the form $(n^r a_{nn})(u_n/n^r)$, we see from Lemma 6.2 and (7.11) that it approaches zero as n becomes infinite. Furthermore the coefficient of u_0 is zero for each n. Hence it follows from (7.13) that

$$\gamma_n - \alpha_n = o(1) + \sum_{k=1}^{n-1} \left\{ a_{nk} - \left(1 - \frac{k}{n}\right)^r \right\} u_k, \tag{7.14}$$

and we may use Lemma 6.1 to obtain

$$\gamma_n - \alpha_n = o(1) + \sum_{k=1}^{n-1} \left(1 - \frac{k}{n}\right)^r \frac{C_{nk}}{n-k} u_k. \tag{7.15}$$

Choosing a constant C such that $|C_{nk}| < C$ when $0 < k < n$, we obtain

$$|\gamma_n - \alpha_n| \leqq o(1) + C \sum_{k=1}^{n-1} \frac{k^{r'}(n-k)^{r'-1}}{n^{r'}} |u_k/k^r|, \tag{7.16}$$

where $r' = \Re(r)$.

* It should be noted that the hypotheses of this theorem are not sufficient to ensure that either of the sequences γ_n or α_n is convergent, and hence that this theorem gives an especially important relation between Cesàro and Riesz transforms.

Now (7.16) shows that (7.12) will follow if $\lim v_n = 0$ implies $\lim V_n = 0$ when V_n is defined by

$$V_n = \sum_{k=1}^{n-1} \frac{k^{r'}(n-k)^{r'-1}}{n^{r'}} v_k. \tag{7.17}$$

Thus we can establish Theorem 7.1 by proving that the transformation defined by (7.17) is regular over the set of sequences which converge to zero. To prove the latter result, it is necessary as well as sufficient* to prove that

$$\lim_{n\to\infty} k^{r'}(n-k)^{r'-1}/n^{r'} = 0 \qquad (k = 1, 2, 3, \cdots), \tag{7.18}$$

and that

$$W_n = \sum_{k=1}^{n-1} k^{r'}(n-k)^{r'-1}/n^{r'} < M \qquad (n = 2, 3, 4, \cdots), \tag{7.19}$$

for some constant M which may depend on r but must be independent of n. It is clear that (7.18) holds for any value of r. That (7.19) holds when $\Re(r) < 0$ follows from (5.20) and (5.21) since $W_n = \Phi_n(n)$. Thus Theorem 7.1 is proved.

8. **Relations between B_r and C_r.** The preceding results enable us to establish the following two theorems.

THEOREM 8.1. *If $\Re(r) < 0$, r not a negative integer, then C_r includes B_r.*

Suppose $\sum u_n$ is summable B_r to L so that $\lim \beta(t) = L$. Then by Theorem 5.1, (5.12) and (5.13) hold and we may use Theorem 7.1 to show that $\lim \gamma_n = L$. Thus Theorem 8.1 is proved.

THEOREM 8.2. *If $-1 < \Re(r) < 0$, then B_r includes C_r; if $\Re(r) \leqq -1$, B_r does not include C_r.*

Suppose $-1 < \Re(r) < 0$ and $\sum u_n$ is summable C_r to L. Then $\lim \gamma_n = L$. Since, as is well known, (7.11) is a necessary condition for summability C_r when $\Re(r) > -1$, we can apply Theorem 7.1 to obtain $\lim \alpha_n = L$; an application of Theorem 5.1 completes the proof of the first part of Theorem 8.2. To prove the second part suppose $\Re(r) \leqq -1$, and, of course, that r is not a negative integer. By Lemma 6.3, there is a series $\sum u_n$ summable C_r for which (5.12) fails; hence by Theorem (5.1), $\sum u_n$ is not summable B_r and the second part of Theorem 8.2 is proved.

Theorems 8.1 and 8.2 yield

---

* Kojima, Tôhoku Mathematical Journal, vol. 12 (1917), pp. 291–326; p. 300.

THEOREM 8.3. *If* $-1<\Re(r)<0$, *then* B_r *and* C_r *are equivalent; if* $\Re(r)\leqq -1$, B_r *and* C_r *are not equivalent.*

THEOREM 8.4. *If* r *is real and* >-1, *then* B_r *and* C_r *are equivalent.*

When $-1<r<0$, this is included in Theorem 8.3. When $r>0$, the result is included in Theorem 4.4.

Cesàro's method C_r of summability is, as is well known, not regular when $\Re(r)<0$. When $-1<\Re(r)<0$, C_r will evaluate only a subset of the set of all convergent series, and will evaluate no divergent series; hence, as might be expected, C_r occupies, for this range of values of r, a prominent place in the theory of series. On the other hand when r is real and <-1, C_r can evaluate to zero certain divergent series of positive terms (see, for example, §6). Owing to this fact, and also to the fact that many useful properties which hold when $\Re(r)>-1$ fail when $\Re(r)\leqq -1$, the method C_r has received little attention when $\Re(r)\leqq -1$.

It is of interest to note that Theorems 8.1 and 8.2 show that B_r is equivalent to C_r over precisely the range of values of r with negative real parts over which C_r has been useful, namely the range $-1<\Re(r)<0$.

In the next section, we will show that summability B_r is significant even when $\Re(r)<-1$.

9. Relations between methods B_r of different orders. In this section we prove six theorems on relations between methods B_r of different orders.

THEOREM 9.1. *If* $\Re(r)<-1$ *and* $\Re(r)\leqq\Re(s)$, *then* B_s *includes* B_r.

Let $\sum u_n$ be summable B_r to L so that $\lim \beta^{(r)}(t)=L$. Then, by Theorem 5.1, $\lim u_n/n^r=0$. We may write

$$\beta^{(r)}(t)-\beta^{(s)}(t)=\sum_{k=0}^{[t]-1}\left\{\left(1-\frac{k}{t}\right)^r-\left(1-\frac{k}{t}\right)^s\right\}u_k$$

$$=\sum_{k=1}^{[t]-1}k^r\left\{\left(1-\frac{k}{t}\right)^r-\left(1-\frac{k}{t}\right)^s\right\}u_k/k^r.$$

We see that Theorem 9.1 will follow if the transformation

$$W(t)=\sum_{k=1}^{[t]-1}k^r\left\{\left(1-\frac{k}{t}\right)^r-\left(1-\frac{k}{t}\right)^s\right\}w_k \tag{9.11}$$

is regular over the set of all sequences w_n which converge to zero.

Letting $d_k(t)$ represent the coefficient of w_k in (9.11), we have evidently

$$\lim_{t\to\infty} d_k(t)=0 \qquad (k=1,2,3,\cdots). \tag{9.12}$$

Also

$$\sum_{k=1}^{[t]-1} |d_k(t)| \leqq \sum_{k=1}^{[t-1]} k^{r'}\left\{\left(1-\frac{k}{t}\right)^{r'}+\left(1-\frac{k}{t}\right)^{s'}\right\} \leqq 2\sum_{k=1}^{[t]-1} k^{r'}\left(1-\frac{k}{t}\right)^{r'}$$

where $r' = \Re(r)$ and $s' = \Re(s)$. Since $r' < -1$,

$$\sum_{k=1}^{[t/2]} k^{r'}\left(1-\frac{k}{t}\right)^{r'} \leqq 2^{-r'}\sum_{k=1}^{[t/2]} k^{r'} < 2^{-r'}\sum_{k=1}^{\infty} k^{r'},$$

and

$$\sum_{k=[t/2]+1}^{[t]-1} k^{r'}\left(1-\frac{k}{t}\right)^{r'} \leqq 2^{-r'}\sum_{k=[t/2]+1}^{[t]-1} (t-k)^{r'} < 2^{-r'}\sum_{k=1}^{\infty} k^{r'}.$$

Hence

$$(9.13) \qquad \sum_{k=1}^{[t]-1} |d_k(t)| < 2^{-r'+2}\sum_{k=1}^{\infty} k^{r'}.$$

The conditions (9.12) and (9.13) ensure that (9.11) has the desired property and Theorem 9.1 is proved.

From Theorem 9.1, we obtain at once

THEOREM 9.2. *If* $\Re(r) = \Re(s) < -1$, *then* B_r *and* B_s *are equivalent.*

From Theorems 9.1 and 5.1 we obtain

THEOREM 9.3. *If* $\Re(r) < -1$ *and* $\sum u_n$ *is summable* B_r *to* L, *then* $\sum u_n$ *converges to* L, *the convergence being absolute.*

That $\sum u_n$ must converge to L follows from the fact that B_0, which includes B_r by Theorem 9.1, represents convergence. Again, by Theorem 5.1, $\lim u_n/n^r = 0$; hence $|u_n| < n^{r'}$, $r' = \Re(r) < -1$, for all sufficiently great n, and absolute convergence of $\sum u_n$ follows. Thus Theorem 9.3 is proved.

THEOREM 9.4. *If* $-1 < \Re(r) < \Re(s) < 0$, *then* B_s *includes* B_r. *If* $-1 < r < s$, *then* B_s *includes* B_r.

The first part of the Theorem follows from the fact (Theorem 8.3) that B_r and C_r are equivalent when $-1 < \Re(r) < 0$ and the fact that C_s includes C_r when $-1 < \Re(r) < \Re(s)$. The second part follows from a similar application of Theorem 8.4.

To complete Theorems 9.1 and 9.4, it would be desirable to determine whether B_s includes B_r when $-1 = \Re(r) < \Re(s) < 0$. Neither the method of proof of Theorem 9.1 nor that of Theorem 9.4 throws light on this question. A partial answer to this question is given by the following theorem.

THEOREM 9.5. *If* $-1=r<\Re(s)<0$, *then* B_s *includes* B_r.

We shall give a proof of Theorem 9.5 after having proved Theorem 10.1 below. After having proved Theorem 9.5, we can use Theorems 9.1, 9.4, and 9.5 to give a relation of inclusion between any two methods B_r of real orders, namely

THEOREM 9.6. *If* $r<s$, *then* B_s *includes* B_r.

10. **Consideration of** A_r. Since it is sometimes convenient to use transformations involving a continuous parameter, and at other times a discontinuous parameter, it is important to know whether A_r and B_r are equivalent, and whether the results which we have established for B_r hold also for A_r.

Using Theorem 8.4 and the result of Riesz that A_r and C_r are equivalent when $-1<r<1$, we see that A_r and B_r are equivalent when $-1<r<1$. We proceed to show that A_r and B_r have very different properties when $\Re(r)<-1$.

A series $\sum u_k$ is said to be summable by the Abel method P to L if $\sum u_k x^k$ converges for $|x|<1$ and $\lim_{x\to 1-}\sum u_k x^k=L$. We shall say that $\sum u_k$ is summable P^* to L if $\sum u_k x^k$ converges for all sufficiently small $|x|$ and generates an analytic function $u(x)$ such that $\lim_{x\to 1-} u(x)=L$.† It is evident that P^* includes P and that P does not include P^*.

THEOREM 10.1. *If* $\Re(r)\leqq -1$ *and* $r\neq -1$, *then* P^* *does not include* A_r; *if* r *is real and* $\geqq -1$, *then* P^* *includes* A_r.

Let $\sum u_k$ be summable A_r to L; then $\alpha_n\to L$ where

$$(10.11)\qquad (n+1)^r\alpha_{n+1}=\sum_{k=0}^{n}(n+1-k)^r u_k.$$

From (10.11) we obtain when $|x|<1$,

$$(10.12)\qquad \sum_{n=0}^{\infty}(n+1)^r\alpha_{n+1}x^n=\sum_{n=0}^{\infty}\sum_{k=0}^{n}(n+1-k)^r u_k x^n.$$

Letting $u(x)$ be the analytic function determined by the equation

$$(10.13)\qquad \sum_{n=0}^{\infty}(n+1)^r\alpha_{n+1}x^n=u(x)\sum_{n=0}^{\infty}(n+1)^r x^n,$$

we see that when $|x|$ is sufficiently small, say $|x|<\delta$, $u(x)$ is a convergent power series in x. A comparison of (10.12) and (10.13) suffices to show that

---

† This modification of Abel's method was introduced by Silverman-Tamarkin, Mathematische Zeitschrift, vol. 29 (1928), pp. 161–170; p. 169.

$$u(x) = \sum_{k=0}^{\infty} u_k x^k, \qquad |x| < \delta; \tag{10.14}$$

hence $u(x)$ is the analytic function generated by $\sum u_k x^k$. That P^* includes A_r when r is real and $\geqq -1$ follows at once from the conditions for regularity† of the transformation defined by (10.13); this also follows from a result of Silverman-Tamarkin, loc.cit.

We shall prove the first part of Theorem 10.1 by a method which shows that P^* and A_r are inconsistent when $\Re(r) < -1$. Corresponding to each complex r, let $\sum u_n^{(r)}$ be the series having for its A_r transform the sequence $\alpha_1 = 1$, $\alpha_n = 0$ when $n > 1$. Then $\sum u_n^{(r)}$ is summable A_r to 0. Using (10.14) and (10.13), we see that the analytic function $u^{(r)}(x)$ generated by $\sum u_k^{(r)} x^k$ is given by

$$u^{(r)}(x) \sum_{n=0}^{\infty} (n+1)^r x^n = 1. \tag{10.15}$$

But when $r = -1 - ih$, h real and $\neq 0$, we have as $x \to 1^-$‡

$$\sum_{n=0}^{\infty} (n+1)^{-1-ih} x^n \sim \Gamma(ih)\{\log(1/x)\}^{ih}$$

so that $\lim_{x\to 1^-} u^{(r)}(x)$ does not exist and $\sum u_n^{(r)}$ is non-summable P^*. On the other hand, if $\Re(r) < -1$, then $\sum (n+1)^r$ converges to $\zeta(-r)$ which is finite and different from zero; hence $\sum u_n^{(r)}$ is summable P^* to $1/\zeta(-r)$ which is finite and different from the A_r value of $\sum u_n^{(r)}$. Thus Theorem 10.1 is proved.

We pass now to a proof of Theorem 9.5. Let $\sum u_k$ be summable B_{-1} to L. Then by Theorem 5.1, $nu_n \to 0$ and $\sum u_k$ is summable A_{-1} to L. Then by Theorem 10.1, $\sum u_k$ is summable P^* to L. But $\sum u_k x^k$ must converge when $|x| < 1$ since $nu_n \to 0$; hence $\sum u_k$ is summable P to L. Therefore, by Tauber's Theorem§ $\sum u_k$ must converge to L. Since $nu_n \to 0$ and $\sum u_k$ converges to L, it follows‖ that $\sum u_k$ is summable C_s for every s with $\Re(s) > -1$. Finally summability B_s for every s with $-1 < \Re(s) < 0$ follows from Theorem 8.3 and Theorem 9.5 is proved.

We have shown in the proof of Theorem 10.1 that when $\Re(r) < -1$, the transformation A_r can evaluate to 0 a series which is not summable P to 0 and which is therefore not convergent to 0. Using this result and Theorem 9.3, we obtain

† W. A. Hurwitz, loc. cit., p. 20.

‡ Lindelöf, *Le Calcul des Résidus*, p. 139.

§ A. Tauber, Monatshefte für Mathematik und Physik, vol. 8 (1897), pp. 273–277.

‖ Hardy and Littlewood, Proceedings of the London Mathematical Society, (2), vol. 11 (1912), p. 462.

THEOREM 10.2. *If* $\Re(r)<-1$, *then* A_r *and* B_r *are not equivalent.*

Theorem 10.1 also shows that the methods A_r do not, in contrast to the methods B_r, form for real values of r a set of consistent methods of summability whose effectiveness increases steadily as r increases.

We can now see that (5.12) is not a consequence of (5.13) when $\Re(r)<-1$ by proving

THEOREM 10.3. *When* $\Re(r)<-1$, *the condition* $u_n/n^r\to 0$ *is not necessary in order that* $\sum u_n$ *may be summable* A_r.

If the condition were necessary, it would follow from Theorem 5.1 that A_r and B_r would be equivalent and Theorem 10.2 would be contradicted.

In the next section, we give a theorem which is interesting in connection with Theorem 10.3, and give further properties of A_r.

11. **Consideration of A_r when $\Re(r)<\zeta$.** Let ζ, $-2<\zeta<-1$, be the real negative root of the equation.

$$2^r+3^r+4^r+\cdots=1. \tag{11.01}$$

We shall now prove

THEOREM 11.1. *If* $\Re(r)<\zeta$ *and* $\sum u_n$ *is bounded* A_r, *then* u_n/n^r *is bounded for all* $n>0$.†

Let $\sum u_n$ be bounded A_r, $\Re(r)<\zeta$, so that α_n, being defined by

$$\alpha_n=\sum_{k=0}^{n-1}\left(1-\frac{k}{n}\right)^r u_k, \tag{11.11}$$

is a bounded sequence. Since $r'=\Re(r)<\zeta$, it follows that

$$0<\theta_r=2^{r'}+3^{r'}+4^{r'}+\cdots<1. \tag{11.12}$$

Choose an index $p>1$ so great that

$$2^{-r'+1}\sum_{k=p}^{\infty}k^{r'}<(1-\theta_r)/2 \tag{11.13}$$

and let a sequence v_n be defined by the formulas $v_n=0$, $n<p$; $v_p=u_0+u_1+\cdots+u_p$; and $v_n=u_n$, $n>p$. Then

$$\alpha_n=o(1)+\sum_{k=p}^{n-1}\left(1-\frac{k}{n}\right)^r v_k=o(1)+\sum_{k=p}^{n-1}k^r\left(1-\frac{k}{n}\right)^r(v_k/k^r).$$

Hence we can prove Theorem 10.1 by showing that boundedness of W_n

† $\sum u_n$ is said to be bounded A_r when its A_r transform is a bounded sequence.

implies boundedness of w_n whenever

$$W_n = \sum_{k=p}^{n} k^r \left(1 - \frac{k}{n+1}\right)^r w_k, \qquad n > p. \tag{11.14}$$

Let d_{nk} represent the coefficient of w_k in (11.14). Then when $n > 2p$,

$$\sum_{k=p}^{n-p-1} | d_{nk} | \leqq 2^{-r'+1} \sum_{k=p}^{\infty} k^{r'} < (1 - \theta_r)/2; \tag{11.15}$$

the first inequality being obtained by considering separately the sums when k ranges from p to $[(n+1)/2]$ and from $[(n+1)/2]+1$ to $n-p-1$. Also

$$\lim_{n\to\infty} \sum_{k=n-p}^{n-1} | d_{nk} | = \sum_{k=2}^{p} k^{r'} < \theta_r. \tag{11.16}$$

Combining (11.15) and (11.16), we obtain

$$\limsup_{n\to\infty} \sum_{k=p}^{n-1} | d_{nk} | \leqq (1 + \theta_r)/2 < 1. \tag{11.17}$$

Since

$$\lim_{n\to\infty} d_{n,n} = 1, \tag{11.18}$$

we may use (11.17) and the fact that $d_{nk} = 0$ when $k < p$ to obtain

$$\liminf_{n\to\infty} \left\{ | d_{n,n} | - \sum_{k=0}^{n-1} | d_{nk} | \right\} > 0. \tag{11.19}$$

Owing to (11.19), the fact that the transformation (11.14) has the desired property results from the following lemma.

LEMMA 11.2. *If the coefficients in the transformation*

$$W_n = \sum_{k=0}^{n} d_{nk} w_k \tag{11.21}$$

satisfy (11.19) *and if W_n is a bounded sequence, then w_n is a bounded sequence.*

To prove this lemma, let w_n be an unbounded sequence; we shall show that W_n is an unbounded sequence. Since w_n is unbounded, we can choose an increasing sequence n_j of indices such that $|w_{n_j}| \geqq |w_k|$ when $0 \leqq k < n_j$. Then

$$\begin{aligned} | W_{n_j} | &\geqq - \sum_{k=0}^{n_j-1} | d_{n_j k} | \, | w_k | + | d_{n_j n_j} | \, | w_{n_j} | \\ &\geqq \left\{ | d_{n_j n_j} | - \sum_{k=0}^{n_j-1} | d_{n_j k} | \right\} | w_{n_j} | . \end{aligned}$$

But lim $|w_{n_j}| = +\infty$ and using (10.19) we see that lim $|W_{n_j}| = +\infty$. Hence W_n is an unbounded sequence, Lemma 11.2 is proved, and Theorem 11.1 follows.

THEOREM 11.2. *If* $\Re(r) < \zeta$, *every series summable* A_r *is convergent, but not necessarily to the value to which it is summable.*

That a series summable A_r must be convergent follows from Theorem 11.1; in fact boundedness A_r is sufficient to ensure absolute convergence of $\sum u_n$. That the A_r and convergence values need not be equal is shown by the series $\sum u_n^{(r)}$ used in the proof of Theorem 10.1.

Since C_r can evaluate certain divergent series when $\Re(r) < -1$, r not a negative integer, and A_r can evaluate only absolutely convergent series when $\Re(r) < \zeta$, it follows that C_r and A_r are not equivalent when $\Re(r) < \zeta$.

The methods A_r, $\Re(r) < \zeta$, may be of use for classification of convergent series; but use of such methods for evaluation of series is open to the objection that they are, by Theorem 11.2, inconsistent with convergence.

12. **Conclusion.** In conclusion we point out that while A_r, B_r, and C_r are not mutually equivalent when $\Re(r) < -1$, there is an important class of series over which these methods are equivalent. In fact, a combination of Theorems 5.1 and 7.1 yields the following theorem.

THEOREM 12.1. *If* $\Re(r) < 0$, *being* $\neq -1, -2, \cdots$ *when* C_r *is involved, and* lim $u_n/n^r = 0$, *and if* $\sum u_n$ *is summable by one of the methods* A_r, B_r, *and* C_r, *then it is summable to the same value by the other two methods.*

BROWN UNIVERSITY,
PROVIDENCE, R. I.

ON SOME FUNCTIONALS*

BY

STANISLAW SAKS†

1. In this paper we intend to give a new proof and various generalizations of the following theorem due to Hahn:‡

If $\{f_n(t)\}$ *is a sequence of summable functions in the interval* $J=(0, 1)$ *and if* $\lim_n \int_E f_n(t)dt$ *exists for every measurable set* $E \subset J$, *then the indefinite integrals* $F_n(x) = \int_0^x f_n(t)dt$ *are equally absolutely continuous in* J *and therefore converge to an absolutely continuous function.*

The proof will be based on a theorem of Baire which has proved useful in many similar cases.§ Incidentally there will be given a generalization of another theorem concerning sequences of functional transformations and published in a previous paper by the author.¶

2. We shall denote by R the space of measurable characteristic functions in the interval $J=(0, 1)$, i.e., functions which almost everywhere assume two values only, 0 and 1. The distance of two functions $x(t)$, $y(t)$ in R is defined by the formula

$$d(x, y) = \|x - y\| = \int_J |x(t) - y(t)| dt, \qquad d(x, 0) = \|x\| = \int_J |x(t)| dt.$$

With this definition of the distance, R is a metric complete space. It is not linear but nevertheless has simple properties which in some cases may replace linearity. We shall state them in the following lemma.

Lemma. (i) *If* u *and* x_0 *belong to* R *and if* $\|u\| \leqq r$, *then there exist in* R *elements* u_1, u_2 *such that*

$$u_1 = u_2 + u, \quad d(x_0, u_1) \leqq r, \quad d(x_0, u_2) \leqq r.$$

(ii) *If* $x_1 \subset R$, $x_2 \subset R$, $d(x_1, x_2) \leqq r$, *then there exist in* R u_1, u_2 *such that*

$$\|u_1\| \leqq r, \quad \|u_2\| \leqq r, \quad x_1 - u_1 \subset R, \quad x_2 = x_1 - u_1 \dotplus u_2.$$

(iii) *If* x *is an element of* R *and* $r>0$, *then there exist a finite number of elements* $u_1, u_2, \cdots, u_n$ *such that* $\|u_i\| \leqq r$ $(i=1, 2, \cdots, n)$ *and*

$$x = u_1 + u_2 + \cdots + u_n.$$

---

* Presented to the Society, October 29, 1932; received by the editors August 24, 1932.

† International Research Fellow.

‡ Hahn [1]. Another proof is given in the book of Banach [2, pp. 152–158]. References in brackets refer at the bibliography at the end of this paper.

§ See for instance Banach and Steinhaus [1], Banach [2].

¶ Saks [1].

To prove (i) we simply put

$$u_1(t) = u(t) + x_0(t)[1 - u(t)],$$
$$u_2(t) = x_0(t)[1 - u(t)].$$

To prove (ii) we put

$$u_1(t) = x_1(t)[1 - x_2(t)],$$
$$u_2(t) = x_2(t)[1 - x_1(t)].$$

To prove (iii) let n be an arbitrary integer $>r^{-1}$. Let v_i ($i=1, 2, \cdots, n$) denote the characteristic function of the interval $((i-1)/n, i/n)$. The functions $u_i(t)=x(t)v_i(t)$ satisfy the condition required by (iii).

The theorem of Hahn will now be stated as follows (in a slightly more general form):

THEOREM 1. *If* $\{f_n(t)\}$ *is a sequence of integrable functions and if the sequence of functionals*

$$F_n(x) = \int_0^1 f_n(t)x(t)dt, \quad x \subset R,$$

defined in the space R, *converges on a set of the second category** $H \subset R$, *then the unctionals* $F_n(x)$ *are equally continuous*† *in* R.

We can suppose that the sequence of the functionals $F_n(x)$ converges to zero for every $x \subset H$; otherwise we could replace the sequence $\{F_n(x)\}$ by the double sequence $\{F_n(x)-F_m(x)\}$. Let ϵ be an arbitrary positive number. Denote by H_n the set of points $x \subset R$ such that $|F_m(x)| \leqq \epsilon/2$ for every $m \geqq n$. Then

$$H \subset \sum_n H_n.$$

As H is by assumption of the second category there exists a value n_0 such that H_{n_0} is also of the second category. On the other hand, by the continuity of the functionals $F_n(x)$, all sets H_n are closed; hence H_{n_0} contains a sphere, say‡ $K=K(x_0; r)$. Let u be an arbitrary element of R such that $\|u\| \leqq r$. Then, by

---

* In the sense of Baire. See for instance Hausdorff, *Mengenlehre*, 1927, pp. 138–145.

† I.e., to every $\epsilon>0$ there corresponds an $\eta>0$ such that $\|x\| \leqq \eta$, $x \subset R$, *implies* $|F_n(x)| \leqq \epsilon$ ($n=1, 2, \cdots$). From the equal continuity of the functionals $F_n(x)$, it readily follows that under the assumptions of Theorem 1 the sequence $\{F_n(x)\}$ converges *everywhere* in R. For a more general result see below, Theorem 4 (i).

‡ In metric spaces, $K(x_0; r)$ will generally denote the sphere whose center is x_0 and radius is r.

the preceding lemma, (i), there exist elements u_1, u_2 in R such that $u_1=u_2+u$, $u_1\subset K$, $u_2\subset K$. Therefore for every $n\geqq n_0$,

$$|F_n(u)| \leqq |F_n(u_1)| + |F_n(u_2)| \leqq \epsilon.$$

Thus the theorem of Hahn is established.

By the same method the following theorem may be proved:

THEOREM 2. *If* $\{f_n(t)\}$ *is a sequence of functions integrable over the interval* $J=(0, 1)$ *and if for all functions* $x(t)$ *of a set of the second category in the space* R

$$\overline{\lim_n}\left|\int_0^1 f_n(t)x(t)dt\right| < \infty,$$

then also

$$\overline{\lim_n}\int_0^1 |f_n(t)|\, dt < \infty .$$

3. In the sequel we shall consider functional transformations defined either in an arbitrary metric complete linear space or in the space R of characteristic functions (§2). The values assumed by these transformations will belong to the space S of all measurable functions defined on a measurable set I. The distance $d(\xi, \eta)$ of two functions $\xi(t)$, $\eta(t)\subset S$ will be defined by the well known formula of Fréchet

$$d(\xi, \eta) = \int_I \frac{|\xi(t) - \eta(t)|}{1 + |\xi(t) - \eta(t)|}\, dt,$$

and we put, as usual, $\|\xi\|=d(\xi, 0)$. With this definition of the distance, $\lim_n d(\xi_n, \xi)=0$ $(\xi_n, \xi\subset S)$ means that the sequence $\{\xi_n(t)\}$ converges in measure to $\xi(t)$.

Let

$$(3.1)\qquad \xi = \xi(x, t) = F(x)$$

be a functional transformation of the kind described above (ξ is a measurable function, x an element of a metric space* E, $t\subset I$). Since E and S are metric spaces it is clear what should be understood by the continuity of the transformation (3.1). This transformation will be called linear if it is continuous and if, for every pair of elements x, y in E,

$$F(x + y) = F(x) + F(y)$$

provided that the sum $x+y$ is defined and belongs to E.†

* For the sake of convenience we shall use the Greek letters ξ, η, $\cdots$ to denote the elements of the space S, i.e., the measurable functions, and italics, x, y, $\cdots$, for the elements of the space where the transformations (3.1) are defined; t will usually denote points of the set I.

† This restriction is necessary in the case of the space R which is not linear.

Functional transformations

$$\xi_n = \xi_n(x, t) = F_n(x) \qquad (n = 1, 2, \cdots) \tag{3.2}$$

($\xi_n \subset S$, $x \subset E$, $t \subset I$) will be called equally continuous in the metric space E if to every $\epsilon > 0$ there corresponds an $\eta > 0$ such that whenever $x', x'' \subset E$ and $d(x', x'') \leqq \eta$, we have

$$\|\xi_n' - \xi_n''\| \leqq \epsilon, \; n = 1, 2, \cdots, \text{where } \xi' = F_n(x'), \; \xi'' = F_n(x'');$$

or, what is equivalent, if to every $\epsilon > 0$ there corresponds an $\eta > 0$ such that $d(x', x'') \leqq \eta$ implies $|\xi_n(x'', t) - \xi_n(x', t)| \leqq \epsilon$ for all $t \subset I$, with the exception at most of a subset* of I of measure less than ϵ.

If A is a measurable subset of I then the number

$$d_A(\xi, \eta) = \int_A \frac{|\xi(t) - \eta(t)|}{1 + |\xi(t) - \eta(t)|} dt = \|\xi - \eta\|_A \qquad (\xi, \eta \subset S)$$

will be called the distance of ξ and η with respect to the set A. With this definition it is clear what is meant by the continuity of a transformation (3.1), or the equal continuity and the convergence of a sequence of transformations (3.2), with respect to a measurable set $A \subset I$.

4. Lemma. *If*

$$\xi_n = \xi_n(x, t) = F_n(x) \qquad (n = 1, 2, \cdots) \tag{4.1}$$

($x \subset E$, $t \subset I$, $\xi_n \subset S$) *is a sequence of continuous functional transformations in a metric complete space E, and if for every element x of a set of the second category $H \subset E$, the inequality*

$$\overline{\lim_n} \, |\xi_n(x, t)| < \infty$$

holds on a set of values of t of measure $\alpha > 0$, then to every ϵ, $0 < \epsilon < \alpha$, there correspond a set $A \subset I$ (independent of x) of measure $\alpha - \epsilon$, a sphere K in E, and a number M, such that

$$|\xi_n(x, t)| \leqq M \qquad (n = 1, 2, \cdots)$$

for every $x \subset K$ and all $t \subset A$ with the exception at most of a set of values of t of measure zero (which might depend on x).

Let $\mathfrak{A}$ be a sequence of measurable subsets of the set I everywhere dense in the space of all measurable subsets† of I. Let $\{A_m\}$ be a sequence of all sets

* This exceptional subset generally depends on n.

† I.e., the set of the characteristic functions of sets of $\mathfrak{A}$ has to be everywhere dense in the space of all measurable characteristic functions defined over I (see §2).

in $\mathfrak{A}$ whose measure is $\geqq \alpha - \epsilon$. Denote by H_m the set of elements $x \subset E$ such that

$$|\xi_n(x, t)| \leqq m \qquad (n = 1, 2, \cdots)$$

almost everywhere in A_m. Then

$$H \subset H_1 + H_2 + \cdots + H_m + \cdots .$$

H is by assumption of the second category. Hence, there exists a number M such that H_M is also of the second category. On the other hand, since the transformations (4.1) are continuous the sets H_m are closed, and therefore H_M contains a sphere K. It is easily seen that the set A_M, the sphere K, and the number M satisfy the required conditions.

THEOREM 3.* *If* $\{\xi_n(x, t) = F_n(x)\}$ $(x \subset E, \xi_n \subset S, t \subset I)$ *is a sequence of linear transformations in a metric complete and linear space E, then there exists a set* $A \subset I$ *such that*

(i) $$\overline{\lim_n}\, |\xi_n(x, t)| < \infty$$

at almost every point $t \subset A$ *and for every* $x \subset E$;

(ii) $$\overline{\lim_n}\, |\xi_n(x, t)| = \infty$$

at almost every point $t \subset I - A$, *and for every* $x \subset E$ *with the exception at most of a set of the first category in E;*

(iii) *the transformations* $F_n(x)$ *are equally continuous with respect to the set* A.†

Let α_0 be the upper bound of all numbers α such that there exists a set $H(\alpha)$ of the second category in E with the property that for every $x \subset H(\alpha)$

$$\overline{\lim_n}\, |\xi_n(x, t)| < \infty \tag{4.2}$$

in a subset of I of measure α.‡ Theorem 3 is trivial when $\alpha_0 = 0$. Hence we may assume $\alpha_0 > 0$. Then, by the preceding lemma, there exists for every p a sphere K_p, a number M_p and a set $A_p \subset I$ of measure $\geqq \alpha_0 - 1/p$ such that for every $x \subset K_p$,

$$|\xi_n(x, t)| \leqq M_p \qquad (n = 1, 2, \cdots), \tag{4.3}$$

---

* This theorem has been proved in our previous paper (Saks [1]) under the assumption that there exist an everywhere dense set E_1 in E such that the sequence $\{F_n(x)\}$ converges for every $x \subset E_1$.

† A may be empty or coincide with the whole set I.

‡ This subset in general depends on x.

almost everywhere on A_p. Let

$$A = \sum_p A_p. \tag{4.4}$$

We shall prove that the set A has properties (i), (ii), (iii) above.

First, the inequality (4.2) holds for almost every $t \subset A_p$ and every $x \subset K_p$, and so, by the linearity of the space E, for almost every $t \subset A_p$ and every $x \subset E$,

$$\overline{\lim_n} \mid \xi_n(x, t) \mid < \infty .$$

Finally, by (4.4) this holds for almost every $t \subset A$ and every $x \subset E$. Hence, property (i) is established.

In order to prove (ii) suppose that there exist a set H of the second category in E such that for every $x \subset H$ the relation (4.2) holds in a subset of $I-A$ of positive measure. Then for every $x \subset H$ this relation would hold in a subset of I of measure $>$meas $A=\alpha_0$, which contradicts the definition of the number α_0.

Finally, let r_p be the radius of the sphere K_p, and let x_0 be an arbitrary element in E such that $\|x_0\| \leqq r_p/(pM_p)$. It easily follows from (4.3) that

$$\mid \xi_n(pM_p x_0, t) \mid \leqq 2M_p \qquad (n = 1, 2, \cdots),$$

and therefore

$$\mid \xi_n(x_0, t) \mid \leqq \frac{2}{p} \qquad (n = 1, 2, \cdots)$$

for almost all $t \subset A_p$, i.e. everywhere in A with the exception at most of a set of measure $\leqq$meas $(A - A_p) \leqq 1/p$. Since p is an arbitrary positive integer, this proves property (iii).

5. The theorem of the preceding section, with the exception of property (iii), may be easily extended to the operations considered in the space R(§2). The proof remains essentially the same, only instead of the linearity of the space E we should use the properties of the space R as stated in the lemma of §2. Property (iii) obviously fails for the space R. However we have the following theorem:

THEOREM 4. *If* $\{\xi_n(x, t) = F_n(x)\}$ $(\xi_n \subset S, t \subset I)$ *is a sequence of linear transformations defined either in a linear metric complete space E or in the space R, then there exists a set* $B \subset I$ *with the following properties*:

(i) *The functions* $F_n(x)$ *are equally continuous and the sequence* $\{F_n(x)\}$ *converges with respect to B for every x in E (or in R).*

(ii) *The sequence* $\{F_n(x)\}$ *diverges with respect to any subset of* $I-B$ *of positive measure for every* x *in* E *(or in* R*) with the exception at most of a set of the first category in* E *(or in* R*).**

We shall prove this theorem for the space R, the proof for the linear spaces being even simpler. We shall use the following lemma which is itself a generalization of the theorem of Hahn of §2.

LEMMA. *If*

$$\xi_n(x, t) = F_n(x) \qquad (n = 1, 2, \cdots; t \subset C \subset I)$$

(where C *is a measurable set) is a sequence of linear transformations in the space* R *converging with respect to* C *on a set of the second category* $H \subset R$*, then* $F_n(x)$ *are equally continuous with respect to* C *(and therefore* $\lim_n F_n(x)$ *is also a linear transformation in* R *with respect to* C*).*

The proof runs exactly as that of Hahn's theorem. We can suppose again that the sequence $\{\xi_n(x,t)=F_n(x)\}$ converges in H to zero. Then, for every $\epsilon>0$, there exist a number n_0 and a sphere K_0 in R, of radius r, such that

$$\|F_n(x)\|_C = \|\xi_n(x, t)\|_C \leqq \epsilon/4$$

for every $x \subset K_0$ and $n \geqq n_0$. From (i) and (ii) of the lemma of §2 it readily follows that $d_C(x_1, x_2) \leqq r$ implies $\|F_n(x_2) - F_n(x_1)\|_C \leqq \epsilon$ for every two points x_1, x_2 in R. This proves our lemma.

We now proceed to the proof of Theorem 4 (for the space R). Denote by α_0 the upper bound of numbers $\alpha \geqq 0$ with the property that there exists a set $B(\alpha) \subset I$ of measure α such that the given sequence $\{F_n(x)\}$ converges with respect to $B(\alpha)$ for all $x \subset R$. Let B be the sum of all sets $B(\alpha)$, $\alpha \leqq \alpha_0$. The set B has both properties (i) and (ii). Indeed, property (i) follows immediately from the definition of B and from the preceding lemma. It remains to establish property (ii). Suppose that for all x of a set of the second category in R, $\{F_n(x)\}$ converges with respect to a subset of $I-B$ of positive measure; this subset depends generally on x, but, by the same argument as in the proof of the lemma of §4 we can determine a fixed set $C \subset I-B$ of positive measure such that $\{F_n(x)\}$ converges with respect to C for all x of a set of the second category $H \subset R$. There exists a sphere K_0 in which H is everywhere dense. By the preceding lemma the transformations $F_n(x)$ are equally continuous with

* Hence, to any sequence of transformations $\{F_n(x)\}$ in a linear space E there correspond two sets A and B (as defined by Theorems 3 and 4). We have obviously $B \subset A$ (with the exception at most of a set of measure zero). In a special case, if there exists an everywhere dense set E_1 in E such that the sequence $\{F_n(x)\}$ converges for every $x \subset E_1$ (at least with respect to the set B), it follows easily from the property (iii) of Theorem 3 that $B=A$ (Saks [1], Theorem 6). This obviously is not true for the space R.

respect to C. Hence, the sequence $\{F_n(x)\}$ converges everywhere in K_0. Then, from parts (ii) and (iii) of the lemma of §2, it easily follows that this sequence converges (with respect to C) everywhere in R. Thus it converges everywhere in R with respect to the set $B+C\subset I$. Since, by assumption, meas $C>0$ and therefore meas $(B+C)>$meas $B=\alpha_0$, this contradicts the definition of α_0. Hence our theorem is proved completely.

6. **Conclusion.** The theorems of §§4 and 5 may be applied to sequences of transformations of the form $F_n(x)=\int_0^1 K_n(s, t)x(s)ds=\xi_n(x, t)$. For instance* if $\{K_n(s, t)\}$ is a sequence of summable functions (or, more generally, summable with respect to s for almost every value of t) and if the sequence

$$\xi_n(x, t) = F_n(x) = \int_0^1 K_n(s, t)x(s)ds \qquad (n = 1, 2, \cdots)$$

converges for all measurable characteristic functions $x(s)$ (or merely on a set of the second category in R) then the operations $F_n(x)$ are equally continuous.† Therefore, if a sequence of integrals $\int_P K_n(s, t)ds$ converges in measure in $(0, 1)$ for every measurable set P and if there exists a function $f(t)$ such that $\int_0^a K_n(s, t)ds$ converges in measure to $f(t)$ for all a, $0\leqq a\leqq 1$, then $\int_P K_n(s, t)ds$ converges in measure to $f(t)$ for all measurable sets in the interval $(0, 1)$.

References

Banach. 1. *Sur la convergence presque partout de fonctionnelles linéaires*, Bulletin des Sciences Mathématiques, vol. 50 (1926), pp. 27–32, 36–43.

Banach. 2. *Teorja operacyj linjowych* (in Polish), Warszawa, 1932.

Banach et Steinhaus. 1. *Sur le principe de la condensation des singularités*, Fundamenta Mathematicae, vol. 9 (1927), pp. 50–61.

Hahn. 1. *Über Folgen linearer Operationen*, Monatshefte für Mathematik und Physik, vol. 32 (1922), pp. 1–88.

Saks. 1. *Sur les fonctionnelles de M. Banach et leur application aux développements des fonctions*, Fundamenta Mathematicae, vol. 10 (1927), pp. 186–196.

---

* Cf. analogous examples in Banach [1].

† If $K_n(s, t)$ reduces to a function of one variable s we find again the theorem of Hahn.

Brown University,
Providence, R. I.

BOOLEAN ALGEBRA. A CORRECTION

BY

EDWARD V. HUNTINGTON

In my paper in these Transactions for January, 1933, the Example 4.5 on page 286 is erroneous, and Postulate 4.5 on page 280 is in fact redundant. Hence the "fourth set" of postulates for Boolean algebra, on the base $(K, +, ')$, should read as follows (the class K being understood to contain at least two distinct elements):

POSTULATE 4.1. *If a and b are in K, then $a+b$ is in K.*

POSTULATE 4.2. *If a is in K, then a' is in K.*

POSTULATE 4.3. $a+b=b+a$.

POSTULATE 4.4. $(a+b)+c=a+(b+c)$.

POSTULATE 4.6. $(a'+b')'+(a'+b)'=a$ [or, $ab+ab'=a$, where, by definition, $ab=(a'+b')'$].

The steps by which the proposition 4.5 ($a+a=a$) is deduced as a theorem from Postulates 4.1, 4.2, 4.3, 4.4, and 4.6, are as follows.*

4.10. $a''=a$. (Proof as on page 281.)

4.11. $a+a'=b+b'$.

Proof (without using 4.5). By 4.6, with 4.3 and 4.4,

$$\begin{aligned} a + a' &= [(a' + b'')' + (a' + b')'] + [(a'' + b'')' + (a'' + b')'] \\ &= [(b' + a'')' + (b' + a')'] + [(b'' + a'')' + (b'' + a')'] = b + b'. \end{aligned}$$

4.12. Definition. $U=a+a'=$the "universe element" of the system.

In particular, $U=U+U'$.

4.15. $a+U'=a$.

By 4.6, 4.10, 4.12,

$$\text{(a)} \qquad U' = (U + U)' + (U + U')' = (U + U)' + U'.$$

By 4.12, (a), 4.4, 4.12,

$$\begin{aligned} U &= U + U' = U + [(U + U)' + U']' \\ &= (U + U') + (U + U)' = U + (U + U)'. \end{aligned}$$

Hence by 4.4, 4.12,

$$\text{(b)} \qquad U + U = U + [U + (U + U)'] = (U + U) + (U + U)' = U.$$

From (a), (b), $U'=U'+U'$, whence by 4.12,

* For an essential step in this proof I am indebted to Mr. B. Notcutt, a Commonwealth Fellow from Oxford University, at present a graduate student in Harvard University. The following article appeared after my brief bibliography was completed: W. V. Quine, *A note on Nicod's postulate*, Mind, vol. 41 (1932), pp. 345–350.

(c) $$(a' + a)' = (a' + a)' + (a' + a)'.$$

By 4.12, 4.6, 4.4,

$$a + U' = a + (a' + a)' = [(a' + a')' + (a' + a)'] + (a' + a)'$$
$$= (a' + a')' + [(a' + a)' + (a' + a)'],$$

whence by (c), 4.6,

$$a + U' = (a' + a')' + (a' + a)' = a.$$

4.5. $a+a=a$.

By 4.15, 4.3, 4.12, 4.6, 4.10,

$$(a + a)' = U' + (a + a)' = (a + a')' + (a + a)' = a'.$$

Hence, by 4.10, $a+a=a$.

The number of postulates in the "fourth set," as thus corrected, is no larger than the number in the "fifth set."

HARVARD UNIVERSITY,
CAMBRIDGE, MASS.

ON A SPECIAL CLASS OF POLYNOMIALS*

BY

OYSTEIN ORE

In the present paper one will find a discussion of the main properties of a special type of polynomials, which I have called p-polynomials. They permit several applications to number theory and to the theory of higher congruences as I intend to show in a later paper, and they also possess several properties which are of interest in themselves.

The p-polynomials are defined in a field with prime characteristic p (modular fields); they form a (usually non-commutative) ring, where ordinary multiplication is replaced by symbolic multiplication, i.e., substitution of one polynomial into another. The p-polynomials are completely characterized by the property that the roots form a modulus. This modulus has a basis, and one shows consequently that the p-polynomials will have a great number of properties in common with differential and difference equations, such that the theory of p-polynomials gives an algebraic analogue to the theory of linear homogeneous differential equations. One finds that the theorems on the representation of differential polynomials will hold also for p-polynomials; the decomposition in symbolic prime factors is not unique, but the factors in two different representations will be similar in pairs. One can introduce the system of multipliers and the adjoint of a p-polynomial and even the Picard-Vessiot group of rationality; it corresponds in this case to a representation of the ordinary Galois group of the p-polynomial by means of matrices in the finite field (mod p). When this representation is reducible, the p-polynomial is symbolically reducible and conversely.

In this paper I have given only the fundamental properties in the theory of p-polynomials; various interesting problems could only be mentioned, while most applications of the theory had to be reserved for another communication. There are a few applications to higher congruences in §5, chapter 1, giving new proofs for theorems by Moore and Dickson; in §6 I give a new and simplified proof for the theorem of Dickson on the complete set of invariants for the linear group (mod p). The invariants are, as one will see, the coefficients of a certain p-polynomial, and a slight generalization of the proof of the fundamental theorem on symmetric functions gives the desired result.

---

* Presented to the Society, February 25, 1933; received by the editors February 3, 1933.

Chapter 1. Properties of p-polynomials

1. Definition of p-polynomials. Let K be an arbitrary field of *characteristic* p where p is a rational prime. In many of the most important applications K is a finite field, but this will not be assumed at this stage.*

A polynomial of the form

$$F_p(x) = a_0x^{p^m} + a_1x^{p^{m-1}} + \cdots + a_{m-1}x^p + a_mx \tag{1}$$

with coefficients in K shall be called a *p-polynomial*; the number m is called the *exponent* of $F_p(x)$. When $a_0=1$, $F_p(x)$ is said to be *reduced*. At times polynomials of the form

$$G_{p^f}(x) = a_0x^{p^{fm}} + a_1x^{p^{f(m-1)}} + \cdots + a_{m-1}x^{p^f} + a_mx \tag{2}$$

will be considered; their properties are quite analogous to those of p-polynomials (1).

The p-polynomials form a modulus, since they are reproduced by addition and subtraction. The pth power of a p-polynomial is again a p-polynomial.

The product of two p-polynomials is not a p-polynomial. It is however fundamental that a new *symbolic multiplication* can be introduced such that the product of two p-polynomials is again a p-polynomial. This multiplication is usually not commutative so that the p-polynomials will form a non-commutative ring.

Let namely

$$G_p(x) = b_0x^{p^n} + b_1x^{p^{n-1}} + \cdots + b_{n-1}x^p + b_nx \tag{3}$$

be a second p-polynomial; we then define the symbolic product $F_p(x)G_p(x)$ as

$$F_p(x) \times G_p(x) = F_p(G_p(x)) \tag{4}$$

and correspondingly $G_p(x)\times F_p(x) = G_p(F_p(x))$. It follows that

$$F_p(x) \times G_p(x) = c_0x^{p^{n+m}} + c_1x^{p^{n+m-1}} + \cdots + c_{n+m-1}x^p + c_{n+m}x, \tag{5}$$

where

$$\begin{aligned} c_0 &= a_1b_0^{p^m}, \\ c_1 &= a_0b_1^{p^m} + a_1b_0^{p^{m-1}}, \\ c_2 &= a_0b_2^{p^m} + a_1b_1^{p^{m-1}} + a_2b_0^{p^{m-2}}, \\ &\cdots\cdots\cdots \quad \cdots\cdots \\ c_{n+m} &= a_mb_n. \end{aligned} \tag{6}$$

* For several of the following theorems it is not even necessary to assume that the coefficient field is commutative.

The exponent of a product is the sum of the exponents of the factors.

One immediately observes, that the theory of p-polynomials is a special case of the theory which I have discussed in the paper *Theory of non-commutative polynomials.** One has only to introduce the correspondence

$$y^0 \to x,\ y \to x^p,\ y^m \to x^{p^m},\ y^{n+m} = y^n y^m \to x^{p^n} x^{p^m} = x^{p^{n+m}}$$

giving in general

$$F_p(x) \to a_0 y^m + \cdots + a_{m-1} y + a_m,$$

to recognize the formal identity of the two theories. Since

$$x^p \times ax = a^p x^p \to ya = a^p y$$

one sees that the two operations *conjugation* and *differentiation* in the general theory are here simply

$$\bar{a} = a^p, \quad a' = 0.$$

From the general theory one can now deduce a great number of facts: In the ring of p-polynomials the symbolic multiplication is associative and distributive with respect to both right-hand and left-hand multiplication. The unit element is $E_p(x) = x$ and there are no divisors of zero, i.e., an identity $A_p(x)B_p(x) = 0$ implies $A_p(x) = 0$ or $B_p(x) = 0$.

A p-polynomial $F_p(x)$ is said to be symbolically right-hand divisible by $D_p(x)$ if $F_p(x) = Q_p(x) \times D_p(x)$. One observes that *when $F_p(x)$ is right-hand symbolically divisible by $D_p(x)$, then $F_p(x)$ is also divisible by $D_p(x)$ in the ordinary sense.* When $F_p(x) = D_p(x) \times Q_p(x)$ we say that $F_p(x)$ is left-hand symbolically divisible by $D_p(x)$.

Let us now consider division for p-polynomials; supposing $m \geqq n$ in (1) and (3) one finds that the differences

$$F_p(x) - a_0 b_0^{-p^{m-n}} x^{p^{m-n}} G_p(x),$$

$$F_p(x) - G_p(x) \times (a_0 b_0^{-1})^{1/p^n} x^{p^{n-m}}$$

do not contain any terms of higher degree than $x^{p^{m-1}}$. It follows, by repetition of this process, that one can write

$$
\begin{aligned}
F_p(x) &= Q_p(x) \times G_p(x) + R_p(x), \\
F_p(x) &= G_p(x) \times P_p(x) + S_p(x),
\end{aligned}
\tag{7}
$$

where the exponents of $R_p(x)$ and $S_p(x)$ are smaller than n. The coefficients

* To appear shortly in the Annals of Mathematics. This paper will be quoted as Ore I.

of $R_p(x)$ are all in K, while the coefficients of $S_p(x)$ lie in some radical field over K.

THEOREM 1. *Symbolic right-hand division of polynomials is always possible, while symbolic left-hand division can only be performed in K, when K is perfect.**

When left-hand divisibility is discussed in the following we shall always assume that K is perfect.

Theorem 1 shows that right-hand (and left-hand) Euclid algorithms exist, and this shows in turn the existence of a unique (reduced) cross-cut $(F_p(x), G_p(x)) = D_p(x)$. When $D_p(x) = x$ we say that $F_p(x)$ and $G_p(x)$ are right-hand symbolically relatively prime, and we can then find such polynomials $A_p(x)$ and $B_p(x)$ of exponents less than m and n respectively that

$$A_p(x) \times F_p(x) + B_p(x) \times G_p(x) = x. \tag{8}$$

We shall finally prove the following theorem:

THEOREM 2. *The symbolical right-hand cross-cut of $F_p(x)$ and $G_p(x)$ is equal to the ordinary cross-cut of these polynomials.*

This follows from our former remark that every symbolic right-hand divisor is also an ordinary divisor of a polynomial and the symbolic Euclid algorithm can therefore also be considered as an ordinary Euclid algorithm.

2. Linear factors. Let us now find the condition that a p-polynomial (1) be divisible symbolically by a linear factor $x^p - \alpha x$. One finds easily

$$F_p(x) = Q_p(x)\,(x^p - \alpha x) + Ax,$$

where

$$A = a_0\alpha^{(p^{m-1}-1)/(p-1)} + a_1\alpha^p + \cdots + a_{m-2}\alpha^{p+1} + a_{m-1}\alpha + a_m. \tag{9}$$

THEOREM 3. *The necessary and sufficient condition that the linear p-polynomial $x^p - \alpha x$ be a symbolic divisor of $F_p(x)$ is that α be a root of*

$$a_0y^{(p^m-1)/(p-1)} + a_1y^{(p^{m-1}-1)/(p-1)} + \cdots + a_{m-2}y^{p+1} + a_{m-1}y + a_m = 0, \tag{10}$$

i.e., α is equal to the $(p-1)$st power of a root of the equation $F_p(x) = 0$.

One can in the same way find the necessary and sufficient condition that $F_p(x)$ be left-hand divisible by $x^p - \alpha x$. The result is, in this case, a little more complicated, namely α must be a root of the equation

$$\begin{aligned} a_0^{1/p^m}y^{(p^m-1)/((p-1)p^{m-1})} &+ a_1^{1/p^{m-1}}y^{(p^{m-1}-1)/((p-1)p^{m-2})} + \cdots \\ &+ a_{m-2}^{1/p^2}y^{(p+1)/p} + a_{m-1}^{1/p}y + a_m = 0. \end{aligned} \tag{11}$$

* Compare Theorem 6, chapter I, Ore I.

From Theorem 3 follows immediately that every p-polynomial will decompose into linear symbolic factors in some finite algebraic extension of K. We shall discuss this decomposition later on.

For the product of linear factors one finds

$$(x^p + a_2x) \times (x^p + a_1x) = x^{p^2} + (a_1^p + a_2)x^p + a_1a_2x,$$

and the following theorem can be proved by induction:

THEOREM 4. *We have*

$$(x^p + a_nx) \times \cdots \times (x^p + a_2x) \times (x^p + a_1x)$$
$$= x^{p^n} + A_1^{(n)}x^{p^{n-1}} + \cdots + A_{n-1}^{(n)}x^p + A_n^{(n)}$$

where

$$A_i^{(n)} = \sum a_{s_1}^{p^{\alpha_1}} a_{s_2}^{p^{\alpha_2}} \cdots a_{s_i}^{p^{\alpha_i}}, \quad s_1 < s_2 < \cdots < s_i,$$

where the sum is to be extended over all s and α such that

$$s_r + \alpha_r = n - i + r.$$

In the simplest case where all a's are equal to one, it is seen that

$$(x^p + x)^{[n]} = x^{p^n} + \binom{n}{1}x^{p^{n-1}} + \binom{n}{2}x^{p^{n-2}} + \cdots + \binom{n}{1}x^p + x.$$

3. The roots of p-polynomials. The roots of p-polynomials have several interesting and characteristic properties. Let us consider an equation

$$F_p(x) = 0; \tag{12}$$

it is obvious that $x=0$ is always a root. Furthermore if ω_1 and ω_2 are roots, it is seen without difficulty that $\omega_1 \pm \omega_2$ are roots.

THEOREM 5. *The roots of an equation* (12) *form a finite modulus.*

When $a_m \neq 0$ we find $F_p'(x) = a_m \neq 0$ and the equation (12) cannot have equal roots. The corresponding modulus must have finite basis and we can state

THEOREM 6. *When $a_m \neq 0$ the roots of $F_p(x) = 0$ form a finite modulus M of rank m. There exists a basis*

$$\omega_1, \omega_2, \cdots, \omega_m \tag{13}$$

for M, such that every root is uniquely representable in the form

$$\omega = k_1\omega_1 + \cdots + k_m\omega_m \quad (k_i = 0, 1, \cdots, p-1). \tag{14}$$

A modulus of the form (14) we shall call a p-modulus. It can be shown that m roots (13) form a basis for M if and only if

$$(15)\qquad \Delta(\omega_1, \omega_2, \cdots, \omega_m) = \begin{vmatrix} \omega_1 & \omega_2 & \cdots \omega_m \\ \omega_1^p & \omega_2^p & \cdots \omega_m^p \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \omega_1^{p^{m-1}} & \omega_2^{p^{m-1}} & \cdots \omega_m^{p^{m-1}} \end{vmatrix}$$

does not vanish.

It should be observed at this point, that if one considers the root of a p^f-equation $G_{p^f}(x)=0$, where $G_{p^f}(x)$ is given by (2), the roots will also form a modulus M and one can find a basis

$$\Omega_1, \Omega_2, \cdots, \Omega_m$$

such that every root is representable in the form

$$\Omega = \kappa_1\Omega_1 + \cdots + \kappa_m\Omega_m$$

where the κ_i run through all the elements of a finite field with p^f elements.

4. **Polynomials with given roots.** We shall next consider the inverse problem: *Given a p-modulus $M_p^{(n)}$ of rank n; to construct a p-polynomial $F(x)$ of exponent n having the elements of $M_p^{(n)}$ for roots.* Let $\omega_1, \cdots, \omega_n$ be a basis for $M_p^{(n)}$. When $n=1$ we find simply

$$(16)\qquad F(x) = x(x-\omega_1)(x-2\omega_1)\cdots(x-(p-1)\omega_1) = x^p - \omega_1^{p-1}x.$$

The general expression can now be found by induction. Let $F_n(x)$ be the p-polynomial having the roots

$$k_1\omega_1 + \cdots + k_{n-1}\omega_{n-1} \qquad (k_i = 0, 1, \cdots, p-1).$$

The elements of $M_p^{(n)}$ will then satisfy the equation

$$F_n(x) = F_{n-1}(x)\cdot F_{n-1}(x-\omega_n)\cdots F_{n-1}(x-(p-1)\omega_n) = 0,$$

and since all occurring polynomials are p-polynomials,

$$F_n(x) = F_{n-1}(x)(F_{n-1}(x) - F_{n-1}(\omega_n))\cdots(F_{n-1}(x) - (p-1)F_{n-1}(\omega_n)),$$

or finally, as in the case $n=1$,

$$(17)\qquad F_n(x) = F_{n-1}(x) - F_{n-1}(\omega_n)^{p-1}F_{n-1}(x),$$

which shows that $F_n(x)$ also is a p-polynomial. Using symbolic multiplication, we can write $F_n(x)$ in the form

$$(18)\qquad F_n(x) = (x^p - F_{n-1}(\omega_n)^{p-1}x) \times F_{n-1}(x).$$

This gives by repeated application

THEOREM 7. *The p-polynomial $F_n(x)$ having the elements of a p-modulus $M_p^{(n)}$ for its roots can be written*

$$(19)\quad F_n(x) = (x^p - F_{n-1}(\omega_n)^{p-1}x) \times (x^p - F_{n-2}(\omega_{n-1})^{p-1}x) \times \cdots \times (x^p - \omega_1^{p-1}x)$$

where $\omega_1, \cdots, \omega_n$ is an arbitrary basis for $M_p^{(n)}$. One has also the formula

$$(20)\qquad F_n(x) = \frac{\Delta(\omega_1, \cdots, \omega_n, x)}{\Delta(\omega_1, \cdots, \omega_n)}$$

where Δ denotes the determinant defined by (15).

It is obvious that the polynomial (20) has $\omega_1, \cdots, \omega_n$ and hence all elements of $M_p^{(n)}$ for its roots.

THEOREM 8. *The necessary and sufficient condition that the roots of a polynomial form a modulus is that the polynomial be a p-polynomial.*

The modulus must be finite, and the field of the coefficients must consequently have the characteristic p. The theorem then follows from Theorems 6 and 7.

5. **Applications to higher congruences.** The results of §4 immediately give various theorems on congruences (mod p).

From the definition of $F_n(x)$ and from (20) follows

$$(21)\qquad \frac{\Delta(\omega_1, \cdots, \omega_n, x)}{\Delta(\omega_1, \cdots, \omega_n)} \equiv \prod_{i=1}^{n}\prod_{k_i=0}^{p-1}(x - (k_1\omega_1 + \cdots + k_n\omega_n)) \qquad (\mathrm{mod}\ p)$$

which is a generalization of well known identities in higher congruences. When one compares the last term in x on both sides one obtains the following generalization of Wilson's theorem:

THEOREM 9. *Let $M_p^{(n)}$ be a finite modulus* (mod p) *and let $\omega_1, \cdots, \omega_n$ be a basis for the modulus; then*

$$(22)\qquad \prod_{\omega}\omega \equiv (-1)^n\Delta(\omega_1, \cdots, \omega_n)^{p-1} \qquad (\mathrm{mod}\ p)$$

where $\omega \neq 0$ runs through all elements of $M_p^{(n)}$.

Let us finally apply the formula (21) to the case of $n-1$ basis elements $\omega_1, \cdots, \omega_{n-1}$ and let us put $x = \omega_n$. This gives

$$\Delta(\omega_1, \cdots, \omega_n) \equiv \Delta(\omega_1, \cdots, \omega_{n-1}) \prod_{i=1}^{n-1}\prod_{k_i=0}^{p-1}(\omega_n + k_{n-1}\omega_{n-1} + \cdots + k_1\omega_1) \qquad (\mathrm{mod}\ p)$$

and we have a simple proof of a theorem by E. H. Moore.*

* E. H. Moore, Bulletin of the American Mathematical Society, vol. 2 (1896), p. 189.

THEOREM 10. *The following identity holds*:

$$(23)\quad \Delta(\omega_1, \cdots, \omega_n) \equiv \prod_{i=1}^{n-1} \prod_{k_{i-1}=0}^{p-1} \cdots \prod_{k_1=0}^{p-1} (\omega_i + k_{i-1}\omega_{i-1} + \cdots + k_1\omega_1) \pmod{p}.$$

It can be stated by saying that $\Delta(\omega_1, \cdots, \omega_n)$ is congruent to the product of all different linear expressions in the ω_i, considering two such expressions equal if they are proportional.

Another result is the following:

Let $H_p(x) = A_p(x) \times B_p(x)$; *then*

$$H_p(x) = \prod_{\omega} (B_p(x) + \omega)$$

where ω *runs through the modulus of all roots of* $A_p(x) = 0$ *and the product sign indicates ordinary multiplication.*

This simple remark contains and generalizes various theorems on higher congruences by Mathieu* and Dickson.†

6. **The invariants of linear groups** (mod p). We shall now consider the symmetric functions of the roots of a p-polynomial. From (20) it follows that the p-polynomial corresponding to given modulus $M_p^{(n)}$ has the form

$$(24)\qquad F_p(x) = x^{p^n} + A_1 x^{p^{n-1}} + \cdots + A_{n-1}x^p + A_n x,$$

where

$$(25)\qquad A_i = (-1)^i \frac{\Delta^{(i)}(\omega_1, \cdots, \omega_n)}{\Delta(\omega_1, \cdots, \omega_n)} \qquad (i = 1, 2, \cdots, n),$$

and where $\Delta^{(i)}(\omega_1, \cdots, \omega_n)$ denotes the minor of the term x^{p^i} in the determinant $\Delta(\omega_1, \cdots, \omega_n, x)$. Every symmetric function of the elements of $M_p^{(n)}$ can therefore be expressed by the rational function (25) of the ω_i.

We shall now consider the inverse problem: When is a rational function $F(x_1, \cdots, x_n)$ a symmetric function of the $p^n - 1$ linear forms

$$(26)\qquad \phi_{k_1 \cdots k_n}(x_1, \cdots, x_n) = k_1 x_1 + \cdots + k_n x_n \quad (k_i = 0, 1, \cdots, p-1),$$

the combination $k_1 = \cdots = k_n = 0$ excluded. We shall prove

THEOREM 11. *The necessary and sufficient condition that* $F(x_1, \cdots, x_n)$ *be a symmetric function of the linear forms* (26) *is that* $F(x_1, \cdots, x_n)$ *be an absolute invariant of the full linear group of order* n (mod p).

When $F(x_1, \cdots, x_n)$ is representable as a symmetric function of the forms

* E. Mathieu, Journal de Mathématiques, (2), vol. 6 (1861), pp. 241–323.

† L. E. Dickson, Bulletin of the American Mathematical Society, vol. 3 (1897), pp. 381–389.

(26) it is representable by the coefficients A_i in (24). From the representation (25) it is easily seen that these coefficients are absolute invariants by all linear substitutions of the ω_i with non-vanishing determinant (mod p).

To prove the converse, let $F(x_1, \cdots, x_n)$ be an absolute invariant; one can assume, without loss of generality, that $F(x_1, \cdots, x_n)$ is integral. If we write

$$F(x_1, \cdots, x_n) = \sum_i B_i(x_2, \cdots, x_n)x_1^{\alpha_i}, \tag{27}$$

the $B(x_2, \cdots, x_n)$ must be absolute invariants of the linear group on the $(n-1)$ variables $x_2, \cdots, x_n$. Let us put

$$\Delta = \prod_{k_2,\cdots,k_n=0}^{p-1} (x_1 + k_2x_2 + \cdots + k_nx_n)^{p-1} = x_1^{p^n - p^{n-1}} + \cdots, \tag{28}$$

where the coefficients of Δ as polynomial in x_1 are also invariants of the group in $n-1$ variables. We can now divide $F(x_1, \cdots, x_n)$ by the powers of Δ and obtain a representation of the form

$$F(x_1, \cdots, x_n) = R_t(x_1)\Delta^t + R_{t-1}(x_1)\Delta^{t-1} + \cdots + R_0(x_1), \tag{29}$$

where the coefficients $R_i(x_1)$ are polynomials of degree smaller than the degree of Δ in x_1 and with coefficients which are invariants in the $n-1$ remaining variables. We shall now show that x_1 does not occur in any $R_i(x_1)$. Let us suppose namely that

$$R_i(x_1) = S_0(x_2, \cdots, x_n) + x_1S_1(x_2, \cdots, x_n) + \cdots . \tag{30}$$

It follows from the representation (28) that Δ is invariant under an arbitrary substitution of the form

$$\begin{aligned} &x_1 \to k_1x_1 + \cdots + k_nx_n, \quad k_1 \neq 0, \\ &x_i \to x_i. \end{aligned} \tag{31}$$

Since the representation (29) is unique, all coefficients in (29) must also be invariant under the substitutions (31). From (30) we obtain, however,

$$R_i(x_1) - S_0(x_2, \cdots, x_n) = x_1K(x_1),$$

and applying all substitutions (31) to this identity we find that the difference $R_i(x_1) - S_0(x_2, \cdots, x_n)$ is divisible by Δ, giving $R_i(x_1) = S_0(x_2, \cdots, x_n)$. This gives the special form

$$F(x_1, \cdots, x_n) = R_t(x_2, \cdots, x_n)\Delta^t + \cdots + R_0(x_2, \cdots, x_n) \tag{32}$$

for the representation (29).

The remaining part of the proof is analogous to the proof for the principal

theorem on symmetric functions. The terms of $F(x_1, \cdots, x_n)$ are arranged in decreasing order as usual in this proof, and we assume that

$$ax_1^{\alpha_1}x_2^{\alpha_2}\cdots x_n^{\alpha_n}, \quad \alpha_1 \geqq \alpha_2 \geqq \cdots \geqq \alpha_n, \tag{33}$$

is the principal term. α_1 is then the highest exponent of any power of x_i which occurs in $F(x_1, \cdots, x_n)$; according to (32) α_1 must be divisible by $p^n - p^{n-1}$; α_2 is the highest power of x_2 contained in the invariant $R_t(x_1, \cdots, x_n)$ and it is therefore by the same reason divisible by $p^{n-1} - p^{n-2}$ etc. It follows that the principal term (33) must have the form

$$ax_1^{t_1(p^n - p^{n-1})}x_2^{t_2(p^{n-1}-p^{n-2})}\cdots x_n^{t_n(p-1)}. \tag{34}$$

The invariant A_i in (25) has the principal term

$$\pm\, x_1^{p^n - p^{n-1}}x_2^{p^{n-1}-p^{n-2}}\cdots x_i^{p^{n-i+1}-p^{n-i}}$$

and the difference

$$F(x_1, \cdots, x_n) - (\pm\, aA_1^{t_1-t_2}A_2^{t_2-t_3}\cdots A_n^{t_n})$$

only contains terms lower than (34) and one obtains a representation of $F(x_1, \cdots, x_n)$ by the A_i through repetition of this process. It also follows that if

$$F(x_1, \cdots, x_n) = R(A_1, \cdots, A_n)$$

is the representation of the integral invariant $F(x_1, \cdots, x_n)$ then the coefficients of R belong to the ring generated by the coefficients of F.

An immediate consequence of this proof is

THEOREM 12. *The polynomials*

$$A_i(x_1, \cdots, x_n) = \frac{\Delta^{(i)}(x_1, \cdots, x_n)}{\Delta(x_1, \cdots, x_n)} \qquad (i = 1, \cdots, n) \tag{35}$$

form a fundamental system for all the absolute invariants of the linear group of n variables (mod p).

A relative invariant of the linear group is an expression $G(x_1, \cdots, x_n)$ which is only multiplied by a power of the substitution determinant by a linear substitution (mod p); $\Delta(x_1, \cdots, x_n)$ is a relative invariant and by multiplying by a suitable power of $\Delta(x_1, \cdots, x_n)$ one obtains a very simple proof of a theorem by Dickson*:

---

* L. E. Dickson, these Transactions, vol. 12 (1911), pp. 75–98.

THEOREM 13. *The polynomials*

$$\Delta(x_1, \cdots, x_n),\ A_i(x_1, \cdots, x_n) \qquad (i = 1, \cdots, n-1) \tag{36}$$

form a fundamental system for all relative invariants of the linear group on n variables (mod p).

The polynomial $A_n(x_1, \cdots, x_n)$ in (36) has been omitted since

$$A_n(x_1, \cdots, x_n) = \Delta(x_1, \cdots, x_n)^{p-1}.$$

Dickson has proved Theorem 13 for the somewhat more general case in which the linear group is supposed to have coefficients in an arbitrary finite field. Our proof holds with slight modifications also for this case. In the same paper Dickson considers the "Formenproblem" of the invariants: i.e., the problem of finding the values of the variables x_i for which the invariants assume prescribed values. From our point of view, this is identical with the problem of solving the equation defined by the corresponding p-polynomial, a problem which has already been discussed at some length.

7. **The resultant.** An important invariant of two p-polynomials $F_p(x)$ and $G_p(x)$ defined by (1) and (3) respectively is the so-called p-resultant $R_p(F_p(x), G_p(x))$. Let

$$\omega_1, \cdots, \omega_m;\ \psi_1, \cdots, \psi_n$$

be the basis elements of the two corresponding p-moduli; the determinant $\Delta(\omega_1, \cdots, \omega_m, \psi_1, \cdots, \psi_n)$ is then according to (22) equal to the product of all possible different linear combinations

$$a_1\omega_1 + \cdots + a_m\omega_m + b_1\psi_1 + \cdots + b_n\psi_n \tag{37}$$

in which not all coefficients vanish, and where two expressions (37) are considered to be equal if one can be obtained from the other through multiplication with a rational integer. We then define the p-resultant of $F_p(x)$ and $G_p(x)$ by putting

$$R_p(F, G) = \frac{\Delta(\omega_1, \cdots, \omega_m, \psi_1, \cdots, \psi_n)}{\Delta(\omega_1, \cdots, \omega_m)\Delta(\psi_1, \cdots, \psi_n)}. \tag{38}$$

This resultant is, we see, the product of the differences of all non-vanishing roots of the two polynomials, considering as before two differences $\omega - \psi$ and $k(\omega - \psi)$ as being equal. It is therefore

$$R_p(F, G) = R\left(\frac{F_p(x)}{x}, \frac{G_p(x)}{x}\right)^{p-1}$$

where R denotes the ordinary resultant. I mention without proof that the p-resultant of $F_p(x)$ and $G_p(x)$ can be represented in the form

$$R_p(F_p, G_p) = \frac{\Delta(F_p(\psi_1), \cdots, F_p(\psi_n))}{\Delta(\psi_1, \cdots, \psi_n)} = \frac{\Delta(G_p(\omega_1), \cdots, G_p(\omega_m))}{\Delta(\omega_1, \cdots, \omega_m)}. \tag{39}$$

One can also find a representation of R_p by means of the coefficients of $F_p(x)$ and $G_p(x)$.

8. **The adjoint of a p-polynomial.** To a given modulus $M_p^{(n)}$ we construct the adjoint modulus $\overline{M}_p^{(n)}$

$$\begin{gathered} \bar{\omega}_1 = (-1)^{n+1}\frac{\Delta(\omega_2, \cdots, \omega_n)}{\Delta(\omega_1, \cdots, \omega_n)}, \quad \bar{\omega}_2 = (-1)^{n+2}\frac{\Delta(\omega_1, \omega_3, \cdots, \omega_n)}{\Delta(\omega_1, \cdots, \omega_n)}, \cdots, \\ \bar{\omega}_n = \frac{\Delta(\omega_1, \cdots, \omega_{n-1})}{\Delta(\omega_1, \cdots, \omega_n)}. \end{gathered} \tag{40}$$

We show simply that these numbers are linearly independent and therefore can be regarded as the basis of a modulus $\overline{M}_p^{(n)}$.

In §4, we have found that the reduced polynomial $F_p(x)$ having the modulus $M_p^{(n)}$ for roots will be left-hand divisible by $x^p - \beta x$, where according to (18) and (20)

$$\beta = F_{n-1}^{-}(\omega_n)^{p-1} = \frac{\Delta(\omega_1, \cdots, \omega_n)^{p-1}}{\Delta(\omega_1, \cdots, \omega_{n-1})^{p-1}} = \frac{1}{\bar{\omega}_n^{p-1}}.$$

By changing the order of the basis elements ω_i of $M_p^{(n)}$ one can deduce in the same way that $F_p(x)$ is left-hand divisible by all factors

$$x^p - \bar{\omega}_i^{-(p-1)}x,$$

and finally also by all factors

$$x^p - \bar{\omega}^{-(p-1)}x,$$

where $\bar{\omega}$ is an arbitrary element of the adjoint modulus $\overline{M}_p^{(n)}$.

Let on the other hand $\bar{\omega}$ be an element such that

$$F_p(x) = (x^p - \bar{\omega}^{-(p-1)}x) \times Q_p(x). \tag{41}$$

The roots of $Q_p(x)$ will form a submodulus $M_p^{(n-1)}$ of $M_p^{(n)}$, and since a basis of $M_p^{(n-1)}$ may be completed to a basis for $M_p^{(n)}$ we see that $\bar{\omega}$ must be an element of $\overline{M}_p^{(n)}$. This leads to the following result which may also be used as a definition of $\overline{M}_p^{(n)}$:

THEOREM 14. *The adjoint modulus $\overline{M}_p^{(n)}$ to $M_p^{(n)}$ consists of all elements $\bar{\omega}$ such that the corresponding p-polynomial $F_p(x)$ to $M_p^{(n)}$ has a decomposition of the form* (41).

We shall express this result in a somewhat different form. From (41) we obtain

$$\bar{\omega}^p F_p(x) = ((x\bar{\omega})^p - \bar{\omega}x) \times Q_p(x) = (x^p - x) \times \bar{\omega}x \times Q_p(x). \tag{42}$$

An element κ such that

$$\kappa F_p(x) = (x^p - x) \times R_p(x) \tag{43}$$

shall be called a *multiplier* of $F_p(x)$. It is obvious that the multipliers form a modulus, and from (42) and Theorem 14 we find

THEOREM 15. *The multipliers of a polynomial $F_p(x)$ form a modulus $N_p^{(n)}$ of rank n which is equal to the modulus of the pth powers of the adjoint modulus $\overline{M}_p^{(n)}$ to the modulus $M_p^{(n)}$ of the roots of $F_p(x)=0$.*

Let us now determine the p-polynomial corresponding to the adjoint modulus $\overline{M}_p^{(n)}$ or to the modulus $N_p^{(n)}$ of the multipliers, which is virtually the same problem. If $F_p(x)$ is left-hand divisible by $x^p-\beta x$, then $\beta=\kappa^{-(p-1)/p}$, where κ is a multiplier, and the condition (11) for left-hand linear factors gives

THEOREM 16. *The multipliers of*

$$F_p(x) = x^{p^n} + A_1 x^{p^{n-1}} + \cdots + A_n x \tag{44}$$

are the roots of the equation

$$\overline{F}_p(x) = (A_n x)^{p^n} + (A_{n-1}x)^{p^{n-1}} + \cdots + (A_1 x)^p + x = 0. \tag{45}$$

Since the roots of $x^p-x=0$ are $0, 1, \cdots, p-1$, we observe the following result: If κ is a multiplier of $F_p(x)$ giving the decomposition (43), then

$$\kappa F_p(x) = \prod_{i=0}^{p-1}(R_p(x) + i) \tag{46}$$

where the product sign denotes ordinary multiplication.

We have in the preceding supposed $F_p(x)$ to be reduced. If $F_p(x)$ in (44) has the highest coefficient A_0 then the multipliers will be $\kappa' A_0^{-1}$, where κ' is a multiplier of the corresponding reduced polynomial.

In general we shall call the polynomial

$$\overline{F}_p(x) = (A_n x)^{p^n} + (A_{n-1}x)^{p^{n-1}} + \cdots + (A_1 x)^p + A_0 x \tag{47}$$

the adjoint of $F_p(x)$. The adjoint of $\overline{F}_p(x)$ is

$$\overline{\overline{F}}_p(x) = A_0^{p^n} x^{p^n} + A_1^{p^n} x^{p^{n-1}} + \cdots + A_n^{p^n} = x^{p^n} \times F_p(x) \times x^{p^{-n}},$$

and for the adjoint of a product one finds

$$\overline{F_p(x) \times G_p(x)} = x^{p^n} \times \overline{G_p(x)} \times x^{p^{-n}} \times \overline{F_p(x)}.$$

It may be more simple to introduce fractional powers and define the adjoint polynomial by putting

$$\overline{F}_p(x) = A_n x + (A_{n-1}x)^{1/p} + \cdots + (A_1 x)^{p^{-n+1}} + (A_0 x)^{p^{-n}}.$$

This expression has the same roots as (47) and it has the simpler properties that the adjoint of a sum is the sum of the adjoints, the adjoint of a product is equal to the product of the adjoints in inverse order, and also simply $\overline{\overline{F}}_p(x) = F_p(x)$.

Let us finally determine when $F_p(x) = \overline{F}_p(x)$, using the definition (47). We obtain the relations

$$A_i^{p^i} = A_i \qquad (i = 0, 1, \cdots, n) \tag{48}$$

and also

$$A_{n-i}^{p^{n-i}} = A_i \qquad (i = 0, 1, \cdots, n),$$

giving

$$A_i^{p^n} = A_i \qquad (i = 0, 1, \cdots, n).$$

THEOREM 17. *When a polynomial $F_p(x)$ is self-adjoint, all coefficients must belong to a finite field of p^n elements; in addition the relations* (48) *must hold.*

CHAPTER 2. FORMAL THEORY

1. The union of p-polynomials. Let $F_p(x)$ and $G_p(x)$ be two p-polynomials given by (1) and (3); the reduced polynomial

$$M_p(x) = [F_p(x), G_p(x)]$$

of smallest degree with coefficients in K which is right-hand symbolically divisible by both $F_p(x)$ and $G_p(x)$ is called the *least common multiple* or the *union* of $F_p(x)$ and $G_p(x)$. From the existence of a Euclid algorithm the existence of the union follows; it has the exponent $m+n-d$, where d is the exponent of the cross-cut $D_p(x) = (F_p(x), G_p(x))$.

Let as before

$$M_p^{(m)} = (\omega_1, \cdots, \omega_m), \quad N_p^{(n)} = (\psi_1, \cdots, \psi_n) \tag{1}$$

be the basis of the two moduli formed by the roots of the two polynomials $F_p(x)$ and $G_p(x)$. The modulus corresponding to $D_p(x)$ is then the modulus formed by the common elements of $M_p^{(m)}$ and $N_p^{(n)}$ while the modulus of the union is

$$T_p^{(n+m-d)} = (\omega_1, \cdots, \omega_m, \psi_1, \cdots, \psi_n).$$

When $F_p(x)$ is relatively prime to $G_p(x)$ we find

$$[F_p(x), G_p(x)] = \frac{\Delta(\omega_1, \cdots, \omega_m, \psi_1, \cdots, \psi_n, x)}{\Delta(\omega_1, \cdots, \omega_m, \psi_1, \cdots, \psi_n)}, \tag{2}$$

because the right-hand side is a reduced polynomial having the same roots as the union. For the same reason we see that also

$$\begin{aligned}[F_p(x), G_p(x)] &= \frac{\Delta(F_p(\psi_1), \cdots, F_p(\psi_n), F_p(x))}{\Delta(F_p(\psi_1), \cdots, F_p(\psi_n))} \\ &= \frac{\Delta(G_p(\omega_1), \cdots, G_p(\omega_m), G_p(x))}{\Delta(G_p(\omega_1), \cdots, G_p(\omega_m))}\end{aligned} \tag{3}$$

represent the union.

As an application let us determine the union of a reduced polynomial $F_p(x)$ and a linear factor $x^p - ax$. Since the roots of the latter are $ka^{1/(p-1)}$ ($k=0, 1, \cdots, p-1$), we find, using formula (17), chapter 1,

$$M_p(x) = F_p(x)^p - F_p(a^{1/(p-1)})^{p-1}F_p(x).$$

If we put

$$\phi(x) = x^{(p^n-1)/(p-1)} + A_1 x^{(p^{n-1}-1)/(p-1)} + \cdots + A_n, \tag{4}$$

a simple reduction shows

THEOREM 1. *The union of a reduced polynomial $F_p(x)$ and a linear polynomial $x^p - ax$ is*

$$\begin{aligned}M_p(x) &= F_p(x) - a\phi(a)^{p-1}F(x) \\ &= x^{p^{n-1}} + (A_1^p - a\phi(a)^{p-1})x + (A_2^p - a\phi(a)^{p-1}A_1)x \\ &\quad + \cdots - a\phi(a)^{p-1}A_n x,\end{aligned} \tag{5}$$

where $\phi(x)$ is defined by (4).

2. **Transformation of p-polynomials.** The existence of a union for two arbitrary p-polynomials permits us to introduce a new operation on p-polynomials, which we shall call *transformation.*

The polynomial

$$A_p^{(1)}(x) = a_0 b_0^{p^{n-d}}[A_p(x), B_p(x)] \times B_p(x)^{-1} = B_p A_p(x) B_p^{-1} \tag{6}$$

is called the transform of $A_p(x)$ by $B_p(x)$. The notation is such that $A_p(x)$ has the exponent n, $B_p(x)$ the exponent m, while d is the exponent of the cross-cut

(7) $D_p(x) = (A_p(x), B_p(x)),\ A_p(x) = \bar{A}_p(x)D_p(x),\ B_p(x) = \bar{B}_p(x)D_p(x).$

Finally, a_0 and b_0 are the highest coefficients of $A_p(x)$ and $B_p(x)$ and the numerical constant in (6) is chosen such that the transform has the same highest coefficient as $A_p(x)$.

When $A_p(x)$ is relatively prime to $B_p(x)$, we say that (6) is a *special transformation*; the transform has then the exponent n. When a cross-cut $D_p(x)$ exists we call the transformation *general*, and the transform has the exponent $n-d$. The general transformation can always be reduced to a special transformation, since it follows from (6) and (7) that

$$(8) \qquad B_pA_p(x)B_p^{-1} = a_0b_0^{p^{n-d}}[\bar{A}_p(x), \bar{B}_p(x)] \times \bar{B}_p(x)^{-1} = \bar{B}_p\bar{A}_p(x)\bar{B}_p^{-1}.$$

When the polynomial $A_p^{(1)}(x)$ is obtained from $A_p(x)$ by a special transformation, we say that $A_p^{(1)}(x)$ is similar to $A_p(x)$. It can be shown that *the notion of similarity is symmetric, reciprocal and associative.*

There exist a large number of results on the transformation of p-polynomials which can all be deduced from the general polynomial theory. They will be given here without proof:*

When $B_p^{(1)}(x) \equiv B_p^{(2)}(x)$ (mod $A_p(x)$) then

$$(9) \qquad B_p^{(1)}A_p(x)(B_p^{(1)})^{-1} = B_p^{(2)}A_p(x)(B_p^{(2)})^{-1}.$$

Furthermore

$$(10) \qquad (C_pB_p)A_p(x)(C_pB_p)^{-1} = C_p(B_pA_p(x)B_p^{-1})C_p^{-1}.$$

From (9) and (10) it follows that if $A_p^{(1)}(x) = B_pA_p(x)B_p^{-1}$, where $B_p(x)$ is relatively prime to $A_p(x)$, then $A_p(x) = B_p^{(1)}A_p^{(1)}(x)(B_p^{(1)})^{-1}$, when $B_p^{(1)}(x)$ is determined such that

$$B_p^{(1)}(x)B_p(x) \equiv x \qquad (\text{mod } A_p(x)),$$

which is always possible according to (8), chapter 1.

For the transformation of a union one finds simply

$$(11) \qquad C_p[A_p(x), B_p(x)]C_p^{-1} = [C_pA_p(x)C_p^{-1}, C_pB_p(x)C_p^{-1}];$$

the corresponding formula does not hold for the cross-cut. For the transform of a product of reduced factors one finds

$$(12) \qquad C_p(B_p(x) \times A_p(x))C_p^{-1} = C_p^{(1)}B_p(x)(C_p^{(1)})^{-1} \times C_pA_p(x)C_p^{-1}$$

where $C_p^{(1)}(x) = A_pC_p(x)A_p^{-1}$. For an arbitrary number of factors one finds

* The proofs follow from Ore I.

a corresponding result, which gives the theorem that the transform of a product is made up of factors which are similar to the factors in the original product.

The following theorem has some important applications in the formal representations of p-polynomials.

If a product $A_p(x)\times B_p(x)$ is divisible by $C_p(x)$ and $C_p(x)$ is relatively prime to $B_p(x)$, then $A_p(x)$ is divisible by $B_pC_p(x)B_p^{-1}$.

Let us now consider the expression for the transform in terms of the roots of the polynomials. From (3) and the definition of transformation follows

THEOREM 2. *When $B_p(x)$ is relatively prime to $A_p(x)$, we have*

$$B_pA_p(x)B_p^{-1} = a_0\frac{\Delta(B_p(\omega_1), \cdots, B_p(\omega_n), x)}{\Delta(B_p(\omega_1), \cdots, B_p(\omega_n))} \tag{13}$$

where the ω_i form a basis for the roots of $A_p(x)$.

When ω is an arbitrary element in the modulus of $A_p(x)$, then the modulus of $B_pA_p(x)B_p^{-1}$ consists of all numbers $B_p(\omega)$ and this holds even in the general case. The transformation is consequently analogous to the Tschirnhausen transformation for algebraic equations.

As an application let us find the transform of a linear polynomial x^p-ax by an arbitrary polynomial $F_p(x)$. From Theorem 1 follows

$$F_p(x^p - ax)F_p^{-1} = x^p - a\phi(a)^{p-1}x. \tag{14}$$

One can easily determine when two linear expressions

$$x^p - ax, \quad x^p - bx \tag{15}$$

are similar. According to (14) every polynomial similar to a linear polynomial can be obtained from it by transformation with an expression cx, and there follows from (14)

THEOREM 3. *Two linear polynomials* (15) *are similar when the quotient $ab^{-1}=c^{p-1}$ is a $(p-1)$st power in K.*

3. **Decomposition into prime factors.** We shall say that two reduced polynomials $A_p(x)$ and $B_p(x)$ are *transmutable* if $A_p(x)$ can be represented in the form

$$A_p(x) = B_pA_p^{(1)}(x)B_p^{-1},$$

where $A_p^{(1)}(x)$ is similar to $A_p(x)$. In this case the product

$$A_p(x)\times B_p(x) = [A_p^{(1)}(x), B_p(x)] = A_p^{(1)}B_p(x)A_p^{(1)}$$

can be written in two ways, such that the factors are similar, but occur in

different order. We shall say that one representation is obtained from the other by *transmutation*. As an example let us find when two linear factors are transmutable. Let

$$A_p(x) = x^p - ax,\ B_p(x) = x^p - bx,\ A_p^{(1)}(x) = x^p - cx;$$

then according to (14)

$$B_p A_p^{(1)}(x) B_p^{-1} = x^p - c(c-b)^{p-1}x$$

and c must be a root of the equation

$$c(c-b)^{p-1} = a.$$

A prime polynomial $P_p(x)$ in K is a polynomial which has no reduced symbolical divisors except itself and x. Every polynomial similar to a prime polynomial is also prime. One can then prove the following theorem*:

THEOREM 4. *Every reduced polynomial has a decomposition into prime factors. Two different decompositions of the same polynomial will have the same number of factors; the factors will be similar in pairs by a suitable ordering, and one decomposition can be obtained from the other through transmutation of factors.*

It is easily seen that one cannot expect the decomposition to be unique; if $F_p(x)$ is an arbitrary polynomial with the exponent n, then $F_p(x)$ is divisible by all p linear factors $x^p - \omega^{p-1}, x$, where ω is an arbitrary root.

4. **Completely reducible polynomials.** We shall say that a polynomial $F_p(x)$ in K is *completely reducible* when it is the union of prime polynomials. It can then be represented by a basis

$$F_p(x) = [P_1(x), \cdots, P_r(x)]$$

where each prime polynomial $P_i(x)$ is relatively prime to the union of the others. We can also show the following:

The necessary and sufficient condition that a polynomial be completely reducible is that two consecutive prime factors in an arbitrary prime polynomial decomposition always be transmutable.

The union of all prime polynomials, which divide an arbitrary polynomial $F_p(x)$ on the right, we shall call the *maximal completely reducible factor* of $F_p(x)$ and denote by $H_p^{(1)}(x)$. Then

$$F_p(x) = F_p^{(1)}(x) \times H_p^{(1)}(x),$$

and $F_p^{(1)}$ can be treated the same way; there follows

* Ore I, Theorem 1, chapter 2.

THEOREM 5. *Every polynomial has a unique representation as product of maximal completely reducible factors.*

From the general theory a large number of results on completely reducible polynomials can be deduced.* We shall however only mention a few facts, which we shall apply at a later point.

We shall say that a completely reducible polynomial is uniform, when it is only divisible by similar prime polynomials. The necessary and sufficient condition that a completely reducible polynomial be uniform is that the basis contain only similar prime polynomials.

Let $F_p(x)$ now be an arbitrary completely reducible polynomial; the union of all prime divisors of $F_p(x)$ which are similar to a given prime polynomial $P_p(x)$, we shall call a maximal uniform component of $F_p(x)$. It then follows that

Every completely reducible polynomial is uniquely representable as the union of maximal uniform components.

Let finally

$$F_p(x) = [P_p^{(1)}(x), \cdots, P_p^{(r)}(x)]$$

be an arbitrary completely reducible polynomial. If $F_p(x)$ is to be divisible by any prime polynomial $P_p(x)$ different from the basis elements, then at least two basis elements must be similar. Any prime divisor of $F_p(x)$ has to be similar to one of the basis elements, and if $P_p^{(1)}(x) = AP_p(x)A^{-1} \neq P_p(x)$ we could have constructed the basis such that $P_p(x)$ and $P_p^{(1)}(x)$ were basis elements. When conversely an arbitrary polynomial $F_p(x)$ is divisible both by $P_p(x)$ and the similar polynomial $P_p^{(1)}(x)$, we see that

$$F_p(x) \equiv 0, \qquad F_p(x) \times A_p(x) \equiv 0 \qquad (\text{mod } P_p(x))$$

and from a theorem in §2, it follows that $F_p(x)$ is also divisible by all polynomials $BP_p(x)B^{-1}$, where $B_p(x)$ is an arbitrary polynomial of the form

$$B_p(x) = k_1x + k_2A_p(x) \quad (k_1, k_2 = 0, 1, \cdots, p-1).$$

Since the roots of $P_p^{(1)}(x)$ are different from those of $P_p(x)$, it is easily seen that $BP_p(x)B^{-1}$ is different from $P_p(x)$ and $P_p^{(1)}(x)$ when $k_1 \neq 0$ and $k_2 \neq 0$. This shows that

The necessary and sufficient condition that a completely reducible polynomial be divisible by a prime polynomial different from those occurring in a basis representation is that the basis representation contain at least two similar prime polynomials.

---

* See Ore I, §2, chapter 2.

One can also state this by saying that *the basis representation of a completely reducible polynomial is unique, when none of the components are similar.*

5. **Decomposable and distributive polynomials.** In Theorem 4 and Theorem 5 we have found two different representations of p-polynomials; several others can be found, but only two other representations of importance will be mentioned briefly.

A polynomial is said to be *decomposable* when there exists a representation

$$F_p(x) = [A_p(x), B_p(x)] \tag{16}$$

where $A_p(x)$ is relatively prime to $B_p(x)$; $F_p(x)$ is said to be *indecomposable* when no such representation exists. We can prove

THEOREM 6. *Every polynomial can be represented as the union of a number of indecomposable polynomials*

$$F_p(x) = [A_p^{(1)}(x), \cdots, A_p^{(r)}(x)], \tag{17}$$

where each indecomposable polynomial $A_p^{(i)}(x)$ is relatively prime to the union of the others; when two or more different representations (17) *exist, they will all have the same number of components, which will be similar in pairs.*

A polynomial $F_p(x)$ shall be said to be *distributive* when there exists a decomposition (16), where $A_p(x)$ and $B_p(x)$ are *proper divisors* of $F_p(x)$; a cross-cut $C_p(x)$ of $A_p(x)$ and $B_p(x)$ may perhaps exist; when no such decomposition (16) exists, we shall say that $F_p(x)$ is *non-distributive.*

For the proofs of the following theorems it is necessary to assume that K *is perfect*; one can then state

THEOREM 7. *The necessary and sufficient condition that a polynomial $F_p(x)$ be non-distributive is that $F_p(x)$ have only a single left-hand prime divisor $P(x)$.*

We shall say that the non-distributive polynomial $F_p(x)$ *belongs to* $P(x)$. It is easily seen that every left-hand divisor of $F_p(x)$ is also non-distributive and belongs to the same prime polynomial $P(x)$. One can also prove

THEOREM 8. *Let the completely reducible polynomial*

$$A_p(x) = [P_1(x), \cdots, P_r(x)] \tag{18}$$

be the union of all prime polynomials dividing a given polynomial $F_p(x)$ on the left. Then every representation of $F_p(x)$ as the union of non-distributive components has the form

$$F_p(x) = [C_1(x), \cdots, C_r(x)], \tag{19}$$

where the non-distributive polynomial $C_i(x)$ belongs to a prime polynomial similar to $P_i(x)$ $(i=1, 2, \cdots, r)$.

We have supposed that (19) is a *shortest* representation; i.e., we have omitted all components which divide the union of the others.

6. **The invariant ring.** We shall now define a certain characteristic group G_I, *the invariant group*, and also a characteristic ring R_I, *the invariant ring*, corresponding to an arbitrary p-polynomial $F_p(x)$. We make the following definition:

The polynomial $I_p(x)$ is said to be an invariant transformer of $F_p(x)$, when $I_pF_p(x)\, I_p^{-1}$ is a divisor of $F_p(x)$.

It is easy to determine the invariant transformers in some simple cases. Let first $F_p(x)=x^p-ax$; it can then be assumed that $I_p(x)=cx$, and from §2 follows

$$I_pF_p(x)I_p^{-1} = x^p - ac^{p-1}x, \quad c \neq 0,$$

giving the values $c=0$ and $c^{p-1}=1$, i.e., $c=0, 1, \cdots, p-1$. Let next $F_p(x) =x^{p^m}$; using the definition of the transform, one easily finds that every polynomial is an invariant transformer.

The definition of the invariant transformers can easily be modified in the following way:

THEOREM 9. *The necessary and sufficient condition that $I_p(x)$ be an invariant transformer of $F_p(x)$ is that*

$$F_p(x) \times I_p(x) \equiv 0 \qquad (\bmod F_p(x)). \tag{20}$$

This condition (20) immediately shows that the sum, difference, and product of two invariant transformers is again an invariant transformer, and the ring of all invariant transformers is the invariant ring of $F_p(x)$.

When an invariant transformer $I_p(x)$ is relatively prime to $F_p(x)$ we must have

$$I_pF_p(x)I_p^{-1} = F_p(x). \tag{21}$$

The invariant transformers satisfying (21) form the invariant group. It is obvious that the product of two such polynomials has the same property, and to show the group property it only remains to show the existence of an inverse. Since $I_p(x)$ is relatively prime to $F_p(x)$, we can determine an $I_p^{(1)}(x)$ such that

$$I_p^{(1)}(x) \times I_p(x) \equiv x \qquad (\bmod F_p(x)),$$

and it is easily seen that also $I_p^{(1)}(x)$ satisfies (21).

Let now α be a root of

$$F_p(x) = 0. \tag{22}$$

from (20) follows that $I_p(\alpha)$ is also a root of (22) for an arbitrary root α and an arbitrary invariant transformer $I_p(x)$. The invariant transformer therefore permutes the roots of (22), or, expressed in a different way, it transforms the modulus formed by the roots of (22) into itself or a submodulus. When all the roots of (22) are different, the invariant transformer $I_p(x)$ is uniquely determined by the transformation it produces, since $I_p(\alpha)=I_p^{(1)}(\alpha)$ for all α implies $I_p(x)\equiv I_p^{(1)}(x)$ (mod $F_p(x)$). Since the number of roots of (22) is finite we obtain

THEOREM 10. *When all the roots of $F_p(x)=0$ are different, the invariant ring and the invariant group are finite.*

When $F_p(x)=0$ has equal roots, then

$$F_p(x) = x^{p^r}G_p(x),$$

and the invariant ring of $F_p(x)$ will be identical with the invariant ring of $G_p(x)$, when considered (mod $G_p(x)$). Incidentally, these remarks also show that the polynomials cx^{p^r} are the only ones for which all the polynomials are invariant transformers.

From the fact that the invariant ring is finite follows that it is an algebra over the finite field (mod p) and the invariant ring has a basis, such that every element can be represented in the form

$$I_p(x) \equiv c_1 I_p^{(1)}(x) + \cdots + c_r I_p^{(r)}(x) \qquad (\text{mod}\, F_p(x))$$

where $c_i=0, 1, \cdots, p-1$. The invariant ring defined here should more specifically be called the right-hand invariant ring. There also exists a left-hand invariant ring having similar properties; for a left-hand invariant transformer $J_p(x)$ one must have as in (20)

$$J_p(x) \times F_p(x) = F_p(x) \times I_p(x), \tag{23}$$

and here $I_p(x)$ must be a right-hand invariant transformer according to definition. When conversely $I_p(x)$ is an invariant right-hand transformer it is easily seen that

$$J_p(x) = F_p(x) \times I_p(x) \times F_p(x)^{-1} \tag{24}$$

is a left-hand invariant transformer of $F_p(x)$.

THEOREM 11. *The left-hand and right-hand invariant rings and groups are directly isomorphic through the correspondence* (24).

Let us finally determine the invariant ring of a prime polynomial $P_p(x)$. In this case every $I_p(x)\not\equiv 0$ (mod $F_p(x)$) has an inverse, and the invariant

ring is a field. Since this field has a finite number of elements, it follows from a theorem of Wedderburn that it is commutative.

THEOREM 12. *The invariant ring of a prime polynomial $P_p(x)$ is a commutative, finite field.*

The invariant ring of a p-polynomial is closely connected with the structure and representations of the given polynomial and several interesting results can be obtained. It will however carry us too far to study these problems here.

CHAPTER 3. CONNECTION BETWEEN p-POLYNOMIALS AND ORDINARY POLYNOMIALS

1. Polynomials belonging to a p-polynomial. We shall finally study some of the connections between p-polynomials and ordinary polynomials in K. First of all we shall show that an arbitrary polynomial $f(x)$ of nth degree always divides a p-polynomial. Let us divide all pth powers of x by $f(x)$; this gives relations of the form

$$x^{p^i} \equiv c_{n-1}^{(i)} x^{n-1} + \cdots + c_0^{(i)} \pmod{f(x)} \quad (i = 0, 1, 2, \cdots). \tag{1}$$

The powers $1, x, x^2, \cdots$ on the right-hand side of the $\nu \leqq n$ first congruences (1) can now be eliminated, and on the left-hand side this gives a p-polynomial $F_p(x)$ with the exponent ν which is divisible by $f(x)$. Since $F_p(x)$ obviously is the p-polynomial with the smallest exponent having this property, it follows from Theorem 2, chapter 1, that every other p-polynomial $\phi_p(x)$ having the same property must be symbolically divisible by $F_p(x)$.

THEOREM 1. *Every polynomial $f(x)$ of degree n belongs to a unique, reduced p-polynomial $F_p(x)$ with exponent $\nu \leqq n$, such that $f(x)$ divides $F_p(x)$ and every other p-polynomial $\phi_p(x)$ divisible by $f(x)$ is symbolically divisible by $F_p(x)$.*

The number ν shall be called the *exponent* of $f(x)$. It is easily seen that one can determine $F_p(x)$, when the p-polynomials corresponding to the irreducible factors of $f(x)$ are known. Let namely

$$f(x) = \phi_1(x)^{e_1} \cdots \phi_r(x)^{e_r} \tag{2}$$

be the prime-function decomposition of $f(x)$; we denote by $g(x)$ the product of all different prime factors of $f(x)$:

$$g(x) = \phi_1(x) \cdots \phi_r(x). \tag{3}$$

When $g(x)$ belongs to $G_p(x)$, then $F_p(x)$ must be symbolically divisible by

$G_p(x)$, and since $F_p(x)$ cannot contain equal factors, except when the last coefficient vanishes, it follows that $F_p(x)$ has the form

$$F_p(x) = x^{p^t} \times G_p(x),$$

where t is the smallest exponent such that p^t exceeds all e_i in (2).

One can consequently assume that the polynomial to be considered has no equal factors and therefore is of the form (3). One finds, that when the irreducible factor $\phi_i(x)$ belongs to $\phi_p^{(i)}(x)$, then $g(x)$ belongs to the union

$$G_p(x) = [\phi_p^{(1)}(x), \cdots, \phi_p^{(r)}(x)].$$

2. **The degrees of the factors.** When the roots of the polynomial $f(x)$ are known, the corresponding p-polynomial $F_p(x)$ can be determined in a different way. Let

$$(4) \qquad f(x) = (x - \theta_1) \cdots (x - \theta_n),$$

and let us assume that all roots are different and non-vanishing. In the field $K(\theta_1, \cdots, \theta_n)$ a linear factor $x-\theta_i$ belongs to $x^p-\theta_i^{p-1}x$, and from the last remarks of §1 we obtain

THEOREM 2. *Let the n different non-vanishing numbers*

$$(5) \qquad \theta_1, \theta_2, \cdots, \theta_n$$

be the roots of a polynomial $f(x)$ in K; then $f(x)$ belongs to

$$(6) \qquad F_p(x) = [x^p - \theta_1^{p-1}x, \cdots, x^p - \theta_n^{p-1}x].$$

It is obvious that the coefficients of $F_p(x)$ belong to K, since they are symmetric functions of the elements (5).

It should be noted that there are always polynomials belonging to an arbitrary p-polynomial $F_p(x)$, for instance $F_p(x)$. There are however not always irreducible polynomials belonging to a given p-polynomial, and consequently there exist p-polynomials without primitive roots, i.e., such that every root of $F_p(x)=0$ satisfies a p-equation with lower exponent. As an example let us take

$$F_p(x) = [x^p - ax, x^p - ab^{p-1}x].$$

$F_p(x)$ is the union of two similar p-polynomials with the exponent 1, and its roots are of the form

$$\theta = k_1a^{1/(p-1)} + k_2ba^{1/(p-1)} \qquad (k_1, k_2 = 0, 1, \cdots, p-1),$$

and θ satisfies the equation with exponent 1

$$x^p - (k_1 + k_2b)^{p-1}ax = 0.$$

It would be an interesting problem to determine the necessary and sufficient condition for the existence of primitive roots.

Let us now suppose that the p-polynomial $F_p(x)$ is generated by an ordinary polynomial $f(x)$ with the roots (5) as indicated in Theorem 2. The roots of $F_p(x)$ are then according to (6)

$$M_p = k_1\theta_1 + \cdots + k_n\theta_n \quad (k_i = 0, 1, \cdots, p-1;\ i = 1, \cdots, n). \tag{7}$$

All factors $g(x)$ of $F_p(x)$ have therefore roots lying in the Galois field $K(\theta_1, \cdots, \theta_n)$ and if N is the degree of this Galois field, it follows that the degree of each factor is a divisor of N. This gives in particular

THEOREM 3. *When $F_p(x)$ is generated by an irreducible Galois polynomial $f(x)$ of degree N, then all factors of $F_p(x)$ have degrees equal to N or a factor of N.*

It is possible that even for an arbitrary p-polynomial $F_p(x)$ the theorem holds that if N is the degree of the maximal factor of $F_p(x)$, then all other factors have degrees equal to N or a factor of N. I have only been able to prove this theorem under certain limiting conditions. It should be observed that Theorem 2 gives a generalization of a well known property of the polynomial $x^{p^f} - x \pmod p$.

3. **The Galois group.** Let $F_p(x)$ be a p-polynomial and $f(x)$ a polynomial belonging to $F_p(x)$; when the roots of $f(x)$ are given by (5), then the roots of $F_p(x)$ form the modulus (7). The following is therefore obvious:

THEOREM 4. *The exponent of $f(x)$ is equal to the rank of the modulus* (7).

Choosing the notation in a suitable manner, one can write the modulus (7) in the reduced form

$$M_p = k_1\theta_1 + \cdots + k_\nu\theta_\nu \quad (k_i = 0, 1, \cdots, p-1;\ i = 1, \cdots, \nu). \tag{8}$$

The equations $F_p(x) = 0$ and $f(x) = 0$ define the same Galois field, as one sees from the representation (8) of the roots. Let G be the Galois group of $f(x)$; any permutation S in G will then produce a substitution on the linear expressions (8), and it is easily seen that two different permutations will produce different substitutions. This shows

THEOREM 5. *When ν is the exponent of the polynomial $f(x)$, then there exists a true representation of the Galois group G of $f(x)$ by means of matrices of rank ν in the finite field* (mod p).

We have in the introduction mentioned the analogy between p-polynomials and differential polynomials. To those who are familiar with the Picard-Vessiot theory of linear homogeneous differential equations, it will be clear that the group of linear substitutions on the expressions (8) correspond-

ing to the Galois group G is the analogue of the group of rationality of a differential equation. One may of course obtain a different representation of G by using a different basis for the roots of $F_p(x)$, but it is easily seen that all such representations are similar.

Almost all theorems on the group of rationality have analogues in the theory of p-polynomials. I shall here only mention two results, analogous to theorems by Loewy on differential equations:

THEOREM 6. *The necessary and sufficient condition that a p-polynomial be reducible in K is that the representation of G be reducible.*

When the representation of G is reducible, one can choose a basis for the modulus of the roots, such that there exists a submodulus G' which is transformed into itself by all substitutions of G. The submodulus G' defines a factor $G_p(x)$ of $F_p(x)$ and since $G_p(x)$ is left unchanged by all substitutions in G it has coefficients in K. When conversely $F_p(x)$ has a symbolic factor $Q_p(x)$ it is clear that a reducible representation of G exists. In a similar way we show

THEOREM 7. *When $F_p(x)$ is decomposable,*

$$F_p(x) = [A_p(x), B_p(x)],$$

then the representation of G is also decomposable and equal to the sum of two representations corresponding to $A_p(x)$ and $B_p(x)$, and conversely.

YALE UNIVERSITY,
NEW HAVEN, CONN.

THE DEGREE AND CLASS OF MULTIPLY TRANSITIVE GROUPS, III*

BY

W. A. MANNING

If a group of substitutions of class u (>3) is more than triply transitive, its degree does not exceed $2u+1$. This is Bochert's Theorem† (reduced one unit) and was the most that could be said in an entirely general way up to the present about the degree of highly transitive groups of given class. It will be proved in this paper that a t-ply transitive group of class $u(>3)$ is of degree $n<6u/5+u/t-t$ if $t>23$. It will also be shown that if $t>4$, $n\leqq 2u$; if $t>5$, $n\leqq 5u/3$; if $t>7$, $n<3u/2$; if $t>11$, $n<4u/3$; if $t>21$, $n<5u/4$.

1. On page 648 of DC2 it is proved that for 4-ply transitive groups $n\leqq 2u+1$. The method there used is now extended to 5-ply transitive groups in the proof of the following theorem.

THEOREM I. *If n is the degree and u the class of a 5-ply transitive group, not alternating or symmetric, $n\leqq 2u$.*

If there is a substitution of order 2 and degree u in the group G, $n\leqq 5u/3$ $<2u-1$ ($u\geqq 6$; DC1, p. 463). Then unless all the substitutions of degree u are of order 3, at least one of them is of order >4. Let us say that

$$S = (abcde\cdots)\cdots,$$

and let $S_1, S_2, \cdots, S_w$ be similar to S and a complete set of conjugates under H, the subgroup of G that fixes a, c, and f, where f is a letter of S not adjacent to a or c in S. Since G is triply transitive we can make

$$S_1 = (a)(f)(c\cdots)\cdots.$$

The condition on the letter f can be satisfied if S is of order >5 by a letter of the first cycle of S, while if S is of order 5 there will be in S a second cycle from which it can be taken, because $u>5$.

We now have, quite as on page 648 of DC2,

* Presented to the Society, August 30, 1932, and June 19, 1933; received by the editors October 3, 1932, and (revised and extended) February 10, 1933.

† Manning, these Transactions, vol. 31 (1929), p. 648. This paper will be referred to as DC2, and the first paper bearing the same title and which appeared in these Transactions, vol. 18 (1917), p. 463, will be called DC1.

$$(1)\quad 3w+\frac{3w(u-1)(u-3)}{n-3}-\frac{2w(u-3)}{n-3}-\frac{w(u-2)(u-3)(u-4)}{(n-3)(n-4)}$$
$$-\frac{2w(u-1)}{n-3}-\frac{w(u-1)(u-2)(u-6)}{(n-3)(n-4)}\geqq wu.$$

Writing x for $n-4$ and k for $u-3$, this is

$$(2)\qquad kx^2-(3k^2+k-4)x+2k^3-8k-6\leqq 0.$$

If $x=2k+1$, (2) becomes $-k^2-2$. If $x=2k+2$, it becomes $2k+2$. Therefore, unless all the substitutions of degree u of G are of order 3, $n<2u$.

If all the substitutions of degree u are of order 3, $S=(abc)(def)\cdots$ and $S_1=(b)(d)\ (a\ \cdots)\ \cdots$. As on page 649 of DC2 we set up

$$(3)\quad 3w+\frac{3w(u-1)(u-3)}{n-3}-\frac{2w(u-3)}{n-3}-\frac{w(u-2)(u-3)(u-4)}{(n-3)(n-4)}$$
$$-\frac{w(u-1)}{n-3}-\frac{w(u-1)(u-2)(u-5)}{(n-3)(n-4)}\geqq wu+\frac{2w(n-u)}{(n-3)(n-4)}.$$

With the same x and k as before, this is

$$(4)\qquad kx^2-(3k^2+2k-4)x+2k^3+k^2-7k-2\leqq 0.$$

If $x=2k+3$, the left member is $4k+10>0$, while if $x=2k+2$, it is $-k^2+k+6$, so that $x<2k+3$, or finally $n\leqq 2u$, and our theorem is proved.

2. We take up next

THEOREM II. *If n is the degree and u (>3) the class of a 6-ply transitive group, $n\leqq 5u/3$.*

In the proof of I the well known fact that there is no 5-ply transitive group of class >3 and <8 was used. That there is no 6-ply transitive group of degree <53 is an immediate consequence of the following two theorems:

A. *If a primitive group of class >3 contains a circular substitution of prime order p (>3), $n\leqq p+2$.**

B. *Let q be an integer $\geqq 2$ and <5; p any prime $>q+1$; then the degree of a primitive group which contains a substitution of order p that displaces pq letters (not including the alternating group) is $\leqq pq+q$.*†

For example, were there such a group of degree 25, with its order necessarily a multiple of $25\cdot 24\cdot 23\cdot 22\cdot 21\cdot 20$, it would contain a substitution of order 11 of degree 11 or 22. The first is impossible by A and the second by B.

By I, the 6-ply transitive group G, if of class $\leqq 26$, is of degree $\leqq 52$, an

* C. Jordan, Bulletin de la Société Mathématique de France, vol. 1 (1873), p. 40.

† Manning, these Transactions, vol. 15 (1909), p. 247.

impossibility by A and B, so that only groups of class >26 need be considered.

The theorem is known to be true if one of the substitutions of degree u is of even order (DC1).

According to Dr. Luther the 6-ply transitive groups of class $u(>3)$ in which there are substitutions of degree $u+e$ of even order $(0<e<u/9)$ have

$$n < \frac{4u}{3} + 4e.^* \tag{5}$$

However, it may be that all the substitutions of degree $\leqq u+e$ are of odd order. Suppose that to be the case.

Let S be a substitution of G of degree u, and of order >3. Then its order, being an odd number, is $\geqq 5$:

$$S = (abcde \cdots j) \cdots (\alpha) \cdots .$$

Among the conjugates of S under G there is a substitution

$$S_1 = (ab\alpha \cdots) \cdots (c) \cdots ,$$

and the complete set of conjugates under H (the subgroup of G that fixes the four letters a, b, c, and α) is $S_1, S_2, \cdots, S_w$. Reasoning as in the proof of I,

$$\sum m_i = 2w + \frac{w(u-3)^2}{n-4}, \tag{6}$$

$$\sum q_i = \sum r_i = w + \frac{w(u-3)}{n-4} + \frac{w(u-3)(u-4)^2}{(n-4)(n-5)}. \tag{7}$$

Now

$$S^{-1}S_i^{-1}SS_i = (c\alpha) \cdots ,$$

a substitution of even order, so that

$$\begin{aligned} 6w &+ \frac{3w(u-3)^2}{n-4} - 2w - \frac{2w(u-3)}{n-4} - \frac{2w(u-3)(u-4)^2}{(n-4)(n-5)} \\ &\geqq w(u+e+1). \end{aligned} \tag{8}$$

Here we replace $n-5$ by x and $u-3$ by k, and have

$$(e+k)x^2 + (e-3k^2+3k)x + 2k(k-1)^2 \leqq 0. \tag{9}$$

If $2k-4e-2$ and $2k-4e-1$ are put for x in (9) we obtain $-2ke+16e^3+12e^2+2e$ and $(k-3e)^2+16e^3-5e^2$, respectively. While the first of these numbers

---

* C. F. Luther, American Journal of Mathematics, vol. 55 (1933), p. 77.

may be negative for large values of k, the second is always positive. Therefore $x<2k-4e-1$, that is,

$$n \leqq 2u - 4e - 3. \tag{10}$$

If S is of order 3,

$$S = (abc)(def) \cdots (\alpha) \cdots,$$

$$S_1 = (ab\alpha) \cdots (c) \cdots,$$

and

$$\sum m_i = 2w + \frac{w(u-3)^2}{n-4} \tag{11}$$

as before. But

$$\sum q_i = \sum r_i = w + \frac{w(u-3)^2(u-4)}{(n-4)(n-5)}. \tag{12}$$

The inequality

$$6w + \frac{3w(u-3)^2}{n-4} - 2w - \frac{2w(u-3)^2(u-4)}{(n-4)(n-5)} \geqq w(u+e+1), \tag{13}$$

with x for $n-5$ and k for $u-3$, reduces to

$$(e+k)x^2 + (e - 3k^2 + k)x + 2k^3 - 2k^2 \leqq 0. \tag{14}$$

If $2k-4e$ and $2k-4e+1$ are put for x in (14) we get $-2ke+16e^3-4e^2$ and $(k-3e+1)^2+16e^3-21e^2+8e-1$, respectively. Hence $x<2k-4e+1$, or $n<2u-4e$, or

$$n \leqq 2u - 4e - 1. \tag{15}$$

Dr. Luther's limit (which for $e=0$ is that of DC1), $4u/3+4e$, increases with e, while (15) decreases with e. It is to be proved from (5) and (15) that $n\leqq 5u/3$, irrespective of the presence or absence of substitutions of order 2. Let E be the integral part of $u/12-1/8$, which is the solution for e of the equation

$$2u - 4e - 1 = \frac{4u}{3} + 4e. \tag{16}$$

Now either all the substitutions of degree $\leqq u+E+1$ of G are of odd order, or one of the substitutions of degree $\leqq u+E+1$ is of order 2. Then if we put $E+1$ for e in (5), we have a valid upper limit for the degree of 6-ply transitive groups of class $u(>3)$. Therefore

$$n < \frac{4u}{3} + 4E + 4$$

$$< \frac{4u}{3} + 4\left(\frac{u}{12} + \frac{7}{8}\right)$$

$$< \frac{5u}{3} + \frac{7}{2}\cdot \tag{17}$$

This is not what we set out to prove, but before proceeding farther, let us make use of it to revise the lower limit, $u>26$, of the class of 6-ply transitive groups under which we have been working. From $n>52$, $5u/3+7/2>52$, and therefore $u\geqq 30$.

To find the limit stated in our theorem, we return to the two cases: (1) At least one substitution of degree u is of order >3. (2) All substitutions of degree u are of order 3.

(1) In this case $n\leqq 2u-4e-3$, and the solution of

$$2u - 4e - 3 = \frac{4u}{3} + 4e \tag{18}$$

is $e=u/12-3/8$. This number is not an integer. Let E be its integral part. Since $E+1<u/9$ for $u\geqq 30$, formula (5) holds good for E and for $E+1$. When there are substitutions of order 2 of degree $\leqq u+E$ in G, $n<4u/3+4E$, but if all the substitutions of degree $\leqq u+E$ are of odd order, $n\leqq 2u-4E-3$, and of the two, the latter gives the higher limit and should be retained. But we saw that $4u/3+4E+4$ was a true limit also. Then of these two formulas, $2u-4E-3$ and $4u/3+4E+4$, we are at liberty to choose the lower.

If $u\equiv r$, mod 12 ($r=5, 6, \cdots, 16$), $E=(u-r)/12$, and the two formulas between which we may choose are $5u/3+r/3-3$ and $5u/3+4-r/3$. One or the other gives $n\leqq 5u/3$ unless $r=10$ or 11. If $r=11$, $n<5u/3+1/3$ is equivalent to $n\leqq 5u/3-1/3$. Now Dr. Luther's limit, so concisely stated, is deduced from the inequality

$$n < 5 + \frac{4(u + e - 4)^2}{3(u - 2e/3 - 4)}\cdot \tag{19}$$

When $r=10$, and therefore $E=(u-10)/12$, we substitute $E+1=u/12+1/6$ for e in (19) and have on reduction

$$n < \frac{5u}{3} - \frac{u}{102} + 0.49 + \frac{18.68}{17u - 74}\cdot \tag{20}$$

If $u=58$, $n<5u/3-0.04$; if $u=46$, $n<5u/3+0.07$; if $u=34$, $n<5u/3+0.21$.

Since $u \geqq 30$, and since our formula (10), for $e=E+1=u/12+1/6$, becomes $5u/3-11/3$, we conclude that $n \leqq 5u/3$ when $r=10$.

(2) Every substitution of degree u is of order 3, and $u \equiv 0$, mod 3. Equating the right hand members of the two formulas

$$n \leqq 2u - 4e - 1 \tag{15}$$

and

$$n \leqq \frac{4u}{3} + 4e - 1, \tag{21}$$

we find $e=u/12$. If $u/12$ is a whole number, (15) becomes $n \leqq 5u/3-1$. Let $u \equiv r$, mod 12 ($r=3$, 6, or 9); $E=(u-r)/12$. The condition $E+1<u/9$ is satisfied. We are at liberty to choose between $n<5u/3+r/3-1$ and $n<5u/3+3-r/3$. For $r=3$ and for $r=9$, $n \leqq 5u/3$. For $r=6$ both formulas are $n<5u/3+1$. Again we have recourse to Dr. Luther's original formula (19) and in it put $e=E+1=u/12+1/2$. It reduces to

$$n < \frac{5u}{3} - \frac{u}{102} + 1.90 + \frac{51.89}{17u-78}. \tag{22}$$

When $u=102$, $n<5u/3+0.94$, and because the sum of the last three terms of (22) decreases as u increases, $n \leqq 5u/3$ for $u \geqq 102$. For $u=90$, 78, 66, 54, and 42, $5u/3+1=151$, 131, 111, 91, and 71, respectively. But with the aid of Theorems A and B it is easy to show that there are no 6-ply transitive groups of these degrees and of class >3. Therefore $n \leqq 5u/3$, for all non-alternating 6-ply transitive groups.

3. We now undertake to prove the following fundamental theorem:

THEOREM III. *If n is the degree and u (>3) is the class of a t-ply transitive group ($t>6$) in which all the substitutions of degree $\leqq u+e$ are of odd order, $n<2u-4e-5t+37$.*

Here $e=0, 1, \cdots$. There is a substitution S of order $\geqq 3$ among the substitutions of degree u of G. Because G is of class u, S is a regular substitution:

$$S = (ab \cdots) \cdots (ef \cdots ijk \cdots) \cdots (\alpha) \cdots,$$

and $a, b, \cdots, j$ are the first $t-4$ letters of S. Since G is of sufficiently high transitivity, it contains a substitution $(k\alpha) \cdots (a)(b) \cdots (j) \cdots$ which transforms S into

$$S_1 = (ab \cdots) \cdots (ef \cdots ij\alpha \cdots) \cdots (k) \cdots.$$

The indicated order of the letters of S and S_1 is to be maintained unchanged. It may be that t and the order of S are so related numerically that $S_1 = \cdots (j\alpha \cdots) \cdots$ or $S_1 = \cdots (\alpha \cdots) \cdots$, in which case nothing can be said about the order of $S^{-1}S_1^{-1}SS_1$; but if S_1 has two or more of the letters $a, b, \cdots, j$ preceding α in its cycle of S_1,

$$S^{-1}S_1^{-1}SS_1 = (k\alpha) \cdots,$$

and $S^{-1}S_1^{-1}SS_1$ fixes all the $t-4$ letters $a, b, \cdots, j$ except perhaps the first two of the cycle of S_1 in which α occurs.

Let us assume for the moment that t is such a number that $S_1 = \cdots (ij\alpha \cdots) \cdots, \quad \cdots (hij\alpha \cdots) \cdots, \quad \cdots,$ or $\cdots (ef \cdots ij\alpha) \cdots$. There is a doubly transitive subgroup H of G that fixes the $t-2$ letters $a, b, \cdots, j, k$, and α. Its degree is $n-t+2$. Under H, S_1 is one of a complete set of w conjugate substitutions, $S_1, S_2, \cdots, S_w$. Let S_i have m_i letters of H in common with S. Of these m_i common letters (of H), let S replace q_i by common letters (of H) and let S_i replace r_i by common letters (of H). Then the degree of $S^{-1}S_i^{-1}SS_i$ does not exceed $4+3m_i-q_i-r_i$. For this substitution displaces at most 4 of the $t-2$ letters fixed by H: k, α, and the first two letters of the cycle of S_i in which α occurs. It may displace all the other common letters, m_i in number. Of the letters of S_i that are new to S and are letters of H, only those $m_i - r_i$ that follow common letters are displaced by $S_i^{-1}SS_i$ and therefore by $S^{-1}S_i^{-1}SS_i$. And a like statement holds for $S^{-1}S_i^{-1}S$.

For use in the succeeding paragraphs we note that S displaces $u-t+3$ letters of H, and has $(u-t+3)(u-t+2)$ ordered pairs of letters of H; S has $u-t+3$ sequences in letters both of which are displaced by H if k ends a cycle of S, but only $u-t+2$ if k does not end its cycle.

Now S_1 displaces $u-t+3$ letters of H. The complete set $S_1, S_2, \cdots, S_w$ displaces $w(u-t+3)$ letters of H, one as often as any other because H is transitive. Therefore

$$\sum m_i = \frac{w(u-t+3)^2}{n-t+2}. \tag{23}$$

In S_1 there are $u-t+3$ sequences of letters of H if $S_1 = (ef \cdots j\alpha) \cdots$. But if α is not the last letter of its cycle, there are $u-t+2$ such sequences. Then the total number of these sequences in the set is $w(u-t+3)$ or $w(u-t+2)$ and each, because H is doubly transitive, occurs $w(u-t+3)/[(n-t+2)\cdot(n-t+1)]$ times, or $w(u-t+2)/[(n-t+2)(n-t+1)]$ times, respectively. Hence

$$\sum q_i = \frac{w(u-t+3)^2(u-t+2)}{(n-t+2)(n-t+1)} \tag{24}$$

or

$$\sum q_i = \frac{w(u-t+2)^2(u-t+3)}{(n-t+2)(n-t+1)}, \tag{25}$$

the first, and larger, value of $\sum q_i$ holding only when $S_1 = \cdots (ef \cdots ij\alpha) \cdots$. There are $(u-t+3)(u-t+2)$ ordered pairs of letters of H in S_1 in both cases. Then

$$\sum r_i = \frac{w(u-t+3)^2(u-t+2)}{(n-t+2)(n-t+1)} \tag{26}$$

or

$$\sum r_i = \frac{w(u-t+3)(u-t+2)^2}{(n-t+2)(n-t+1)}, \tag{27}$$

in the two cases respectively.

The substitution $S^{-1}S_1^{-1}SS_1$ is of even order and therefore is by hypothesis of degree $>u+e$. Then

$$4w + \sum(3m_i - q_i - r_i) \geqq w(u+e+1), \tag{28}$$

or, if to $\sum q_i$ and $\sum r_i$ are given their smaller values,

$$4 + \frac{3(u-t+3)^2}{n-t+2} - \frac{2(u-t+3)(u-t+2)^2}{(n-t+2)(n-t+1)} \geqq u+e+1. \tag{29}$$

If $S_1 = \cdots (j\alpha \cdots) \cdots (k) \cdots$, $(j\alpha \cdots)$ is not the first cycle of S_1, while there certainly is in G a substitution T_1, a transform of S by $(a)(b) \cdots (h)(i\alpha)(j\beta)(k\gamma) \cdots$, where β and γ are two letters fixed by S:

$$T_1 = \cdots (ef \cdots gh\alpha) \cdots (i)(j)(k) \cdots .$$

The substitution

$$S^{-1}T_1^{-1}ST_1 = (\alpha i)(ef) \cdots .$$

T_1 is one of w (probably numerically different from the former w) conjugates under H. There are $u-t+5$ letters of H in T_i; in the set $T_1, T_2, \cdots, T_w$ each occurs $w(u-t+5)/(n-t+2)$ times. Then

$$\sum m_i = \frac{w(u-t+5)(u-t+3)}{n-t+2}. \tag{30}$$

In T_i there are $u-t+5$ sequences with both letters displaced by H. Therefore

$$\sum q_i = \frac{w(u-t+5)(u-t+3)(u-t+2)}{(n-t+2)(n-t+1)}. \tag{31}$$

Also

$$\sum r_i = \frac{w(u-t+5)(u-t+4)(u-t+2)}{(n-t+2)(n-t+1)}. \tag{32}$$

Combining,

$$\begin{aligned} 4 &+ \frac{3(u-t+5)(u-t+3)}{n-t+2} - \frac{(u-t+5)(u-t+3)(u-t+2)}{(n-t+2)(n-t+1)} \\ &- \frac{(u-t+5)(u-t+4)(u-t+2)}{(n-t+2)(n-t+1)} \geqq u+e+1. \end{aligned} \tag{33}$$

In particular, this inequality (33) arises when $t=8$ and S is of order 3.

There remains the case in which $S_1 = \cdots (\alpha \cdots) \cdots$, and in which we actually use the transform of S by $(a)(b) \cdots (i)(j\alpha)(k\beta) \cdots$:

$$T_1 = \cdots (\cdots hi\alpha) \cdots (j)(k) \cdots .$$

We say that T_1 is one of a complete set of w conjugates under H. T_1 displaces $u-t+4$ letters of H, and therefore

$$\sum m_i = \frac{w(u-t+4)(u-t+3)}{n-t+2}. \tag{34}$$

In T_i there are $u-t+4$ sequences in letters of H. Therefore

$$\sum q_i = \sum r_i = \frac{w(u-t+4)(u-t+3)(u-t+2)}{(n-t+2)(n-t+1)}. \tag{35}$$

Hence

$$\begin{aligned} 4 &+ \frac{3(u-t+4)(u-t+3)}{n-t+2} \\ &- \frac{2(u-t+4)(u-t+3)(u-t+2)}{(n-t+2)(n-t+1)} \geqq u+e+1. \end{aligned} \tag{36}$$

In the three inequalities (29), (33), and (36), we put $x=n-t+1$, $k=u-3$, and $s=t-7$, and simplify. They become, in order,

$$\begin{aligned} (e+k)x^2 &+ [e-3k^2+(6s+7)k-3s^2-6s-3]x+2k^3-(6s+10)k^2 \\ &+ (6s^2+20s+16)k-2s^3-10s^2-16s-8 \leqq 0, \end{aligned} \tag{37}$$

$$\begin{aligned} (e+k)x^2 &+ [e-3k^2+(6s+1)k-3s^2+3]x+2k^3-(6s+3)k^2 \\ &+ (6s^2+6s-3)k-2s^3-3s^2+3s+2 \leqq 0, \end{aligned} \tag{38}$$

$$\begin{aligned} (e+k)x^2 &+ [e-3k^2+(6s+4)k-3s^2-3s]x+2k^3-(6s+6)k^2 \\ &+ (6s^2+12s+4)k-2s^3-6s^2-4s \leqq 0. \end{aligned} \tag{39}$$

In (37) we put $x=2k-4e-6s-3$, and write the result:

$$(40)\qquad \begin{aligned}&(k-7e-8s-1)^2+16e^3+(48s-29)e^2\\&\qquad\qquad+(48s^2-58s+4)e+16s^3-29s^2+4s.\end{aligned}$$

This is clearly positive for $s\geqq 2$. If $s=1$, it is positive if $e\geqq 1$, and when $e=0$ it reduces to $(u-12)^2-9$, which is positive for $t=8$ and $u\geqq 30$. A similar detailed examination shows it to be positive for all $s\geqq 0$ and $e\geqq 0$. When $x=2k-4e-6s-4$, (37) becomes

$$(41)\qquad \begin{aligned}&-10ke-10sk-2k+16e^3+(48s+28)e^2\\&\qquad\qquad+(48s^2+66s+24)e+16s^3+38s^2+26s+4,\end{aligned}$$

which is negative for some sets of values of u, t, and e. Therefore $x<2k-4e-6s-3$, or

$$(42)\qquad n<2u-4e-5t+32\quad(t>6).$$

We next put $x=2k-4e-6s+2$ in (38) and find

$$(43)\qquad \begin{aligned}&\left(k-5e-6s+\frac{9}{2}\right)^2+16e^3+(48s-45)e^2\\&\qquad\qquad+(48s^2-90s+39)e+16s^3-45s^2+39s-\frac{49}{4}.\end{aligned}$$

This is positive for $s\geqq 2$, in fact, is positive for all $s\geqq 0$, $e\geqq 0$. The value of (38) when $x=2k-4e-6s+1$ is

$$(44)\qquad \begin{aligned}&-6ke-6sk+5k+16e^3+(48s-12)e^2\\&\qquad\qquad+(48s^2-18s-10)e+16s^3-6s^2-15s+5.\end{aligned}$$

Therefore

$$(45)\qquad n<2u-4e-5t+37\quad(t>6).$$

Finally, we seek the limit given by (39). Put $x=2k-4e-6s-1$. Then the left member of (39) becomes

$$(46)\qquad (k-5e-6s)^2+k+16e^3+(48s-21)e^2+(48s^2-42s)e+16s^3-21s^2-s,$$

and this too is positive for $s\geqq 0$; $e\geqq 0$. The result of substituting $2k-4e-6s-2$ for x in (39) is

$$(47)\qquad -6ke-6sk-k+16e^3+(48s+12)e^2+(48s^2+30s+2)e+16s^3+18s^2+2s.$$

The limit on n, deduced from (39), is

$$(48)\qquad n<2u-4e-5t+34\quad(t>6).$$

Of the three results (42), (45), and (48), (45) is to be used when $t>7$. If $t=7$, only (37) and (39) are applicable, the latter when S is of order 3, and therefore the limit for $t=7$ is $n<2u-4e-1$.

4. From this point on only groups more than 7-ply transitive will be in question. If Theorems A, B, and II are applied to 8-ply transitive groups, one quickly finds that $n>158$, and $u>94$. We are now to prove

THEOREM IV. *The degree of an* 8-*ply transitive group of class* $u(>3)$ *is less than* $3u/2$.

Dr. Luther* has proved a remarkable theorem which may be stated as follows:

C. *Let* G *be a more than* $2^a+p_1+p_2+\cdots+p_r$ *times transitive group* ($a\geqq 2$; $p_1, p_2, \cdots, p_r$ *distinct odd primes*) *of class* $u(>3)$; *if* G *contains a substitution of even order of degree* $u+e$,

$$n < u + \frac{2u}{2^a p_1 p_2 \cdots p_r - 2} + e + \frac{2e}{2^a - 2} + 1. \tag{49}$$

If G is 8-ply transitive and includes a substitution of order 2 and degree $\leqq u+e$, this theorem asserts that

$$n < \frac{6u}{5} + 2e + 1, \tag{50}$$

without any restriction upon e. Theorem III states that if G contains no substitution of even order of degree $\leqq u+e$,

$$n < 2u - 4e - 3. \tag{51}$$

Equate the right hand members of these two inequalities and solve for e:

$$e = \frac{2u}{15} - \frac{2}{3}. \tag{52}$$

We have a true upper limit for the degree of all 8-ply transitive groups that are not alternating or symmetric if we put $2u/15+1/3$ for e in (50). Thus

$$\begin{aligned} n &< \frac{22u}{15} + \frac{5}{3} \\ &= \frac{3u}{2} - \frac{u-50}{30} \\ &< \frac{3u}{2}. \end{aligned} \tag{53}$$

* American Journal of Mathematics, vol. 50 (1933).

The last step follows because $u>94$.

5. The next theorem can be disposed of very briefly.

THEOREM V. *If a group of class* $u(>3)$ *is more than* 11-*ply transitive, its degree is less than* $4u/3$.

Let $a=3$, $r=1$, and $p_1=3$ in (49). Then, if there are substitutions of order 2 and of degree $\leqq u+e$ in G,

$$n<\frac{12u}{11}+\frac{4e}{3}+1. \tag{54}$$

Also, for $t=12$, III becomes, all the substitutions of degree $\leqq u+e$ being of odd order,

$$n<2u-4e-23. \tag{55}$$

If

$$\frac{12u}{11}+\frac{4e}{3}+1=2u-4e-23, \tag{56}$$

$e=15u/88-9/2$. Then, as before, when we put $15u/88-7/2$ for e in (54),

$$\begin{aligned} n&<\frac{29u}{22}-\frac{11}{3}\\ &<\frac{4u}{3}. \end{aligned} \tag{57}$$

6. It is possible to go one step farther in the elaboration of these extremely concise limit formulas.

THEOREM VI. *The degree of a t-ply* $(t>21)$ *transitive group of class* $u(>3)$ *is less than* $5u/4-t$.

From C, granting its hypothesis, and putting $a=4$, $p_1=5$, it follows that

$$n<\frac{40u}{39}+\frac{8e}{7}+1. \tag{58}$$

By III, if its hypothesis is granted,

$$n<2u-4e-5t+37. \tag{45}$$

Equating, solving for e, and proceeding as before,

$$(59)\qquad \begin{aligned} n &< \frac{436u}{351} - \frac{10t}{9} + \frac{71}{7} \\ &< \frac{5u}{4} - \frac{10t}{9} - \frac{11u}{1404} + \frac{71}{7} \\ &< \frac{5u}{4} - \frac{10t}{9} - \frac{11}{1404}(u - 1295). \end{aligned}$$

From (59) it is clear than $n<5u/4$ $(t>21)$, so that if $u<1295$, $n<1619$. Since A, B, and a short list of prime numbers tell us that there is no non-alternating 22-ply transitive group of degree <1619,

$$(60)\qquad n < \frac{5u}{4} - \frac{10t}{9}$$

$$(61)\qquad < \frac{5u}{4} - t.$$

7. In what follows it is of advantage to know that the class of a 24-ply transitive group exceeds 7600. This is a consequence of VI if n exceeds 9500, which fact can easily be verified by means of A and B and a list of primes.

THEOREM VII. *Let n be the degree of a t-ply $(t>23)$ transitive group of class u (>3); then*

$$(62)\qquad n < \frac{6u}{5} + \frac{u}{t} - t.$$

Let $s=p_1+p_2+\cdots+p_r$ be the sum of r distinct odd primes, given in advance, and let p be their product. G is a t-ply transitive group, and t is large. Now let a be the largest integer such that

$$2^a < t - s \leqq 2^{a+1}.$$

The solution for e of

$$(63)\qquad \frac{2^a pu}{2^a p - 2} + \frac{2^a e}{2^a - 2} + 1 = 2u - 4e - 5t + 37,$$

where (49) has been set equal to (45), is

$$(64)\qquad e = \frac{(2^a - 2)(2^a p - 4)u}{(5\cdot 2^a - 8)(2^a p - 2)} - \frac{(2^a - 2)(5t - 36)}{5\cdot 2^a - 8}\,.$$

Then, on the insertion of $e+1$ in (49),

$$n < \frac{2^a(6\cdot 2^a p - 8p - 4)u}{(5\cdot 2^a - 8)(2^a p - 2)} - \frac{2^a(5t - 36)}{5\cdot 2^a - 8} + \frac{2^a}{2^a - 2} + 1 \tag{65}$$

$$< \frac{6u}{5} + \frac{hu}{t} - t + j, \tag{66}$$

where

$$h = \frac{(8p + 40)t}{5(5\cdot 2^a p - 8p - 10)} \tag{67}$$

and

$$j = \frac{46\cdot 2^a - 8t - 16}{5\cdot 2^a - 8} + \frac{2}{2^a - 2}\cdot \tag{68}$$

Now let $p=35$, $s=12$. In this case, because t exceeds $12+2^a$, $j<38/5$; and because of $t\leqq 12+2^{a+1}$,

$$h = \frac{64t}{5(35\cdot 2^a - 58)} \tag{69}$$

$$\leqq \frac{128(2^a + 6)}{5(35\cdot 2^a - 58)}. \tag{70}$$

If $a\geqq 7$, we conclude from (70) that $h<4/5$. Then for groups that are more than 140 times transitive,

$$n < \frac{6u}{5} + \frac{u}{t} - t - \frac{1}{5}\left(\frac{u}{t} - 38\right). \tag{71}$$

Dr. Luther has proved in a simple way that for non-alternating t-ply transitive groups

$$n \geqq \frac{t^2 - 2t}{4}. \tag{72}$$

Now for $t>21$, we know that $n<5u/4-t$, or $u>4(n+t)/5$, by VI. Therefore

$$u > \frac{t^2 + 2t}{5}\cdot \tag{73}$$

This inequality (73) holds for all primitive non-alternating t-ply transitive groups, as can be easily seen by examining it for $t=1, 2, \cdots, 21$.

In (71), $u/t>38$ if $(t+2)/5\geqq 38$, that is, if $t\geqq 188$. We have therefore proved (62) in case $t>187$.

Let (66) be written thus:

$$n < \frac{6u}{5} + \frac{u}{t} - t - \frac{1-h}{t}(u - l), \tag{74}$$

where $l = jt/(1-h)$. It is clear that (62) is true when $l < 7600$. Now let

$$\begin{aligned}
140 < t &\leqq 187, & p &= 35; & h &< 0.6, & j &< 7.6, & l &< 3560;\\
95 < t &\leqq 140, & p &= 1001; & h &< 0.8, & j &< 7.0, & l &< 4900;\\
76 < t &\leqq 95, & p &= 35; & h &< 0.6, & j &< 7.6, & l &< 1820;\\
44 < t &\leqq 76, & p &= 35; & h &< 0.92, & j &< 7.6, & l &< 7250;\\
37 < t &\leqq 44, & p &= 5; & h &< 0.94, & j &< 7.7, & l &< 5650;\\
30 < t &\leqq 37, & p &= 33; & h &< 0.96, & j &< 6.7, & l &< 6200;\\
24 < t &\leqq 30, & p &= 15; & h &< 0.90, & j &< 7.4, & l &< 2230;\\
t &= 24, & p &= 7; & h &< 0.94, & j &< 7.5, & l &< 3000.
\end{aligned}$$

It is proved that $n < 6u/5 + u/t - t$ if $t > 23$ and $u > 3$.

STANFORD UNIVERSITY,
PALO ALTO, CALIF.

THE ARITHMETICAL THEORY OF LINEAR RECURRING SERIES*

BY

MORGAN WARD

I. Introduction. The difference equation of order one

1. Let m be an integer greater than one, and let

$$(u): \qquad u_0, u_1, u_2, \cdots, u_n, \cdots$$

be an arithmetical series† of order k; that is, a particular solution of the linear difference equation

$$\Omega_{n+k} = c_1\Omega_{n+k-1} + c_2\Omega_{n+k-2} + \cdots + c_k\Omega_n \tag{1.1}$$

where $c_1, c_2, \cdots, c_k$ and the k initial values $u_0, u_1, \cdots, u_{k-1}$ of (u) are given integers. Then if a_n is the least positive residue of u_n modulo m, we may associate with (u) a second sequence

$$(a): \qquad a_0, a_1, a_2, \cdots, a_n, \cdots$$

which we call the reduced sequence corresponding to (u) modulo m.

It is easily seen that after a finite number of terms, the sequence (a) repeats itself periodically, and that any one of its periods is a multiple of a certain least period which is called the *characteristic number* of (u) (or (a)) modulo m‡. The number of non-repeating terms in (a) is called the *numeric* of (u) modulo m; if it is zero, (u) is said to be *purely periodic*§ modulo m. If all the terms of (u) after a certain point are divisible by m, so that the repeating part of (a) consists of the single residue zero, (u) is said to be a *null sequence* modulo m.

Three important problems immediately suggest themselves: first, to determine the characteristic number and numeric of the sequence (u) as

* Presented to the Society, August 31, 1932; received by the editors September 6, 1932.

† The literature prior to 1917 is summarized in Dickson's *History*, vol. I, chapter XVII. Among the more recent papers, D. H. Lehmer, Annals of Mathematics, (2), vol. 31 (1930), pp. 419–449, treats the case $k=2$, and the author, these Transactions, vol. 33 (1931), pp. 153–165, the case $k=3$. For general k, see R. D. Carmichael, Quarterly Journal of Mathematics, vol. 48 (1920), pp. 343–372. Certain of Carmichael's results were extended by the use of ideals by H. T. Engstrom, these Transactions, vol. 33 (1931), pp. 210–218. I shall refer to these papers by the authors' name and page number. For the bearing of the problem upon elementary number theory, see R. D. Carmichael, American Mathematical Monthly, vol. 36 (1929), pp. 132–143.

‡ This term is due to Carmichael, p. 345.

§ This is always the case if m is prime to c_k in (1.1).

functions of the $2k+1$ integers $c_1, \cdots, c_k, u_0, \cdots, u_{k-1}$ and m*; secondly, given (1.1) and m, to determine least upper bounds for the characteristic number and numeric of any solution of (1.1); and thirdly, given m and k, to determine the least upper bounds for the characteristic number and numeric of any arithmetical series of order k. The bearing of these problems upon the arithmetical properties of such series is evident; nevertheless none of them has as yet been completely solved.†

2. The course of the investigation may best be explained by considering the special case of a difference equation of order one,

$$\Omega_{n+1} = c\Omega_n. \tag{2.1}$$

Any solution (u) of (2.1) is of the form

$$u_n = u_0c^n$$

where u_0 is an integer. It is possible to express this solution as the sum of two other solutions $v_n = v_0c^n$, and $w_n = w_0c^n$ where for the modulus m, (v) is a null sequence with the same numeric as (u), and (w) is a purely periodic sequence with the same characteristic number. The numbers v_0 and w_0 may be determined as soon as u_0 is known.

It readily follows that the numeric and characteristic number of the sequence (u) modulo m are respectively the least values of n such that

$$v_0c^n \equiv 0 \pmod{m}, \quad w_0(c^n - 1) \equiv 0 \pmod{m}. \tag{2.2}$$

In the special case when m is a prime p and w_0 is not divisible by p, the least value of n for which the second of these congruences is satisfied is simply the exponent to which c belongs modulo p. A complete solution of our fundamental problems is thus at present out of the question even for a difference equation of order one. Nevertheless it is of considerable interest to reduce the general problem to its basic constituents. A short analysis discloses that in order to determine the minimal values of n in (2.2) it is sufficient to know

(i) the decomposition of m, v_0, w_0 and c into their prime factors;
(ii) the least value of n such that

$$c^n \equiv 1 \pmod{p}$$

for every prime factor p of m;
(iii) if λ is the least value of n satisfying (ii), the highest power of p dividing $c^\lambda - 1$.

---

* Compare Carmichael, pp. 345, 346.

† Compare Engstrom, p. 218.

Furthermore, (i) alone suffices for the determination of the numeric of (u), and (i) and (ii) alone for the determination of the characteristic number of (u) for all square-free integers m. (ii) is the unsolved problem of determining the exponent to which a given integer belongs for a given prime modulus, while (iii) is equivalent to the (unsolved) problem of the quotients of Fermat: to find the highest power of p dividing $c^{p-1}-1$.

Let us pass now to the general case of a difference equation of order k. Let

$$F(x) = x^k - c_1x^{k-1} - \cdots - c_k$$

denote the polynomial associated with the difference equation (1.1), and (u) as before any solution of (1.1). Then we can associate with (1.1) and m two congruences analogous to (2.2):

$$V(x)x^n \equiv 0 \pmod{m, F(x)}, \qquad W(x)(x^n - 1) \equiv 0 \pmod{m, F(x)},$$

where $V(x)$ and $W(x)$ are two polynomials whose coefficients may be determined as soon as the k initial values of (u) are known. The numeric and characteristic number of (u) modulo m are respectively the least values of n such that the first and second of these congruences are satisfied.

The central result of this investigation is that these minimal values of n may be determined in general provided that we know the following:

[i] (a) the decomposition of m into its prime factors;

(b) the Schönemann decompositions* of $F(x)$, $V(x)$ and $W(x)$ modulo p^N, where p is a prime factor of m;

[ii] for every prime factor p of m and every irreducible polynomial factor $\phi(x)$ of $F(x)$ to the modulus p, the least value of n such that

$$x^n \equiv 1 \pmod{p, \phi(x)};$$

[iii] if λ is the least value of n satisfying [ii], the polynomial $L(x)$ defined by

$$x^\lambda - 1 \equiv pL(x) \pmod{p^2, \phi^2(x)}.$$

We have then a complete analogy with the case of a difference equation of order one. Corresponding to (ii), [ii] is the unsolved problem of determining the period of a mark in a Galois field, while [iii] is a kind of generalization of the problem of the quotients of Fermat.†

The methods employed are elementary in the sense that no use is made either of the theory of ideals or the "fundamental theorem of algebra." Instead free use is made of polynomial congruences to single and double moduli in the spirit of Kronecker's theory of algebraic fields. The difficulties in the algebraic treatment due to discriminantal divisors are thereby evaded.‡

---

* See Fricke's *Algebra*, vol. 2, Braunschweig, 1928, chapter 2, and §7 of the present paper.

† Compare Ward, p. 161.

‡ Compare Engstrom, p. 211.

3. We shall adopt the following terminology in this paper. The term polynomial is restricted to mean a polynomial with integral coefficients; if the leading coefficient of the polynomial is unity, it will be said to be primary. We designate polynomials by $A(x)$, $B(x)$, $\cdots$, $U(x)$, $V(x)$, $\cdots$, $\theta(x)$, $\phi(x)$, $\cdots$. A polynomial is said to be divisible by an integer m when and only when all of its coefficients are divisible by m. The notations Res $\{A(x), B(x)\}$ and $(a, b, \cdots)$ will be used for the resultant of two polynomials $A(x)$ and $B(x)$ and the greatest common divisor of two or more integers $a, b, \cdots$.

If (a) is the reduced sequence corresponding to the solution (u) of (1.1) modulo m, and if μ is a period of (a), we shall say that (u) admits the period μ (modd $m, F(x)$) where it will be recalled that $F(x)=x^k- \cdots -c_k$ is the polynomial associated with the difference equation (1.1). In like manner, we shall refer to the characteristic number of (u) as its characteristic number (modd $m, F(x)$) whenever it is necessary to bring m and $F(x)$ in evidence. The notation

$$(u) \equiv (v),\ (u) \equiv (a) \pmod{m},\ 0 \leqq a < m,$$

is self-explanatory.

The following convenient definition was introduced by H. T. Engstrom*: A number π is said to be a general period of the difference equation (1.1) for the modulus m if every sequence of rational integers (u) satisfying (1.1) has the period π. Let τ be the least such general period for the modulus m. Then it is easily seen that every other general period is a multiple of τ, and that the characteristic number of any particular sequence (u) is a divisor of τ. We shall call τ the principal period of the difference equation (1.1) (modd m, $F(x)$). It possesses the following important property:

THEOREM 3.1. *There exist solutions of* (1.1) *whose characteristic number modulo m is the principal period of* (1.1).

Let (u) and (w) be any two solutions of (1.1). Then if we can determine integers $b_1, b_2, \cdots, b_k$ such that

$$u_n \equiv b_1 w_n + b_2 w_{n+1} + \cdots + b_k w_{n+k-1} \pmod{m},\ n = 0, 1, \cdots,$$

the characteristic number of (w) will be a period of (u). Owing to the linearity of (1.1) these congruences will hold for every n provided that they hold for $n=0, 1, 2, \cdots, k-1$. But a sufficient condition that the k congruences

$$\begin{aligned} b_1 w_0 \quad &+ \cdots + b_k w_{k-1} \equiv u_0, \\ &\vdots \\ b_1 w_{k-1} &+ \cdots + b_k w_{2k-1} \equiv u_{k-1} \end{aligned} \pmod{m}$$

* Engstrom, p. 210.

have integral solutions $b_1, \cdots, b_k$ is that their determinant be prime to m. For that particular sequence (w) with the initial values $w_0 = w_1 = \cdots = w_{k-2} = 0$, $w_{k-1} = 1$, this determinant has the value $(-1)^k$.

Hence the characteristic number of (w) is a general period of (1.1). But the characteristic number of (w) must divide the principal period. Hence it is equal to it.

Thus the principal period is the least upper bound of the characteristic numbers of all solutions of (1.1), and the determination of the characteristic number of (w) gives the solution of the second fundamental problem mentioned in the introduction.

Corollary. *If* (u) *is any solution of* (1.1) *and if* $\Delta(u)$ *denotes the determinant*

$$\Delta(u) = \begin{vmatrix} u_0, & u_1, \cdots, u_{k-1} \\ u_1, & u_2, \cdots, u_k \\ \cdot & \cdot \quad \cdot \\ \cdot & \cdot \quad \cdot \\ \cdot & \cdot \quad \cdot \\ u_{k-1}, & u_k, \cdots, u_{2k-1} \end{vmatrix},$$

then if $\Delta(u)$ *is prime to* m, *the characteristic number of* (u) *is the principal period of* (1.1).

As an application of this corollary, consider the solution (s) of (1.1) with the initial values $s_0 = k$, $s_1 = c_1$, $s_2 = c_1^2 + 2c_2$ and so on, so that if the discriminant of $F(x)$ does not vanish, s_n is the familiar sum of the nth powers of the roots of $F(x) = 0$. It is well known that $\Delta(s)$ equals the discriminant of $F(x)$. Hence *the characteristic number of* (s) *is the principal period of* (1.1) *provided that* m *is prime to the discriminant of* $F(x)$.

II. The relationship with the ring associated with the double modulus

4. We begin by considering the solutions of (1.1) from a group-theoretic stand-point. If we regard any two solutions (u) and (v) of (1.1) as one-rowed matrices we may define their "sum" to be the sequence $(u+v)$:

$$(u) + (v) = (u + v).$$

The set of all solutions of (1.1) form an infinite Abelian group with respect to the operation of vector addition just defined, the identity element of the group being the sequence

$$(0): \qquad 0, 0, \cdots, 0, \cdots.$$

Denote this group by $\mathfrak{U}$ and the corresponding finite group of the reduced sequences (a) by $\mathfrak{A}$. The relationship between these two groups may be conveniently symbolized by writing

$$\mathfrak{U} \equiv \mathfrak{A} \pmod{m}.$$

Now the method of attack upon the fundamental problems mentioned in the introduction is *to set up an isomorphism between the group $\mathfrak{A}$ and the ring of residue classes associated with the double modulus m and $F(x)$*. The problems considered are thus transformed into problems belonging to the theory of congruences to a double modulus which admit of perfectly definite answers.

To set up this isomorphism, it is necessary to define the "product" of two sequences (u) and (v). How this may be done will be explained in §6; for the present, we will confine ourselves to developing the idea of addition of sequences.

THEOREM 4.1. *Every sequence (u) may be uniquely represented modulo m as the sum of a null sequence and a purely periodic sequence with the same numeric and characteristic number.*

Let λ and μ be respectively the numeric and characteristic number of (u) modulo m, and suppose that $\lambda \equiv -r \pmod{\mu}$, where $0 \leqq r < \mu$, so that $\lambda + r = q\mu$.

Set $v_n = u_{\lambda+r+n}$, $w_n = u_n - v_n$ $(n = 0, 1, \cdots)$.

Then (v) is a purely periodic sequence with the characteristic number μ modulo m, and

$$(u) = (v) + (w).$$

(w) is a null sequence modulo m with the numeric λ. For if $n \geqq 0$,

$$\begin{aligned} w_{n+\lambda} &= u_{n+\lambda} - v_{n+\lambda} = u_{n+\lambda} - u_{q\mu+n+\lambda} \equiv 0, \\ w_{\lambda-1} &= u_{\lambda-1} - v_{\lambda-1} = u_{\lambda-1} - u_{q\mu+\lambda-1} \not\equiv 0 \pmod{m}. \end{aligned}$$

Such a representation of (u) is unique modulo m; for if there were a second one

$$(u) = (v') + (w')$$

we would have $(w - w') = (v' - v)$, so that $(w - w')$ would be a purely periodic null sequence. Hence $(w - w') \equiv (0) \pmod{m}$, $(w) \equiv (w')$, $(v) \equiv (v') \pmod{m}$.

It is evident that the set of all null sequences of $\mathfrak{A}$ and the set of all purely periodic sequences of $\mathfrak{A}$ are both sub-groups of $\mathfrak{A}$. If we denote these sub-groups by $\mathfrak{N}$ and $\mathfrak{P}$, we have from Theorem 4.1

THEOREM 4.2. *The group $\mathfrak{A}$ is the direct sum of $\mathfrak{N}$ and $\mathfrak{P}$, where $\mathfrak{N}$ is the group of all null sequences of $\mathfrak{A}$, and $\mathfrak{P}$ is the group of all purely periodic sequences of $\mathfrak{A}$.*

5. If we form from the first n terms of any solution (u) of (1.1) a polynomial of degree $n-1$ in the indeterminate x

$$U_n(x) = u_0x^{n-1} + u_1x^{n-2} + \cdots + u_{n-1},$$

it is easily verified that we have identically in x

$$\begin{aligned} F(x)U_n(x) = x^n\{&u_0x^{k-1} + (u_1 - c_1u_0)x^{k-2} + \cdots + (u_{k-1} - c_1u_{k-2} \\ &- \cdots - c_{k-1}u_0)\} - \{u_nx^{k-1} + (u_{n+1} - c_1u_n)x^{k-2} + \cdots \\ &+ (u_{n+k-1} - c_1u_{n+k-2} - \cdots - c_{k-1}u_n)\}. \end{aligned}$$

Denote the two polynomials in brackets by $U(x)$ and $U^{(n)}(x)$ respectively. Then on considering the identity modulo m, we obtain the congruence

$$x^nU(x) - U^{(n)}(x) \equiv 0 \quad (\text{modd } m, F(x)). \tag{5.1}$$

Assume first that (u) is purely periodic modulo m and admits the period n. Then $U^{(n)}(x) \equiv U(x)$ (mod m), so that (5.1) becomes

$$(x^n - 1)U(x) \equiv 0 \quad (\text{modd } m, F(x)).$$

Conversely if for some n this latter congruence holds, (u) is purely periodic modulo m and admits the period n.

Secondly, assume that (u) is a null sequence modulo m of numeric $\leqq n$. Then $U^{(n)}(x) \equiv 0$ (mod m) and (5.1) becomes

$$x^nU(x) \equiv 0 \quad (\text{modd } m, F(x)).$$

Conversely if for some n this latter congruence holds, (u) is a null sequence of numeric $\leqq n$. We have thus established the following two basic theorems:

FUNDAMENTAL THEOREM ON PURELY PERIODIC SEQUENCES. *If (u) is any solution of the difference equation* (1.1), *then a necessary and sufficient condition that (u) should be purely periodic and admit the period n* (modd m, $F(x)$) *is that*

$$(x^n - 1)U(x) \equiv 0 \quad (\text{modd } m, F(x)), \tag{5.2}$$

where

$$U(x) = u_0x^{k-1} + (u_1 - c_1u_0)x^{k-2} + \cdots + (u_{k-1} - c_1u_{k-2} - \cdots - c_{k-1}) \tag{5.3}$$

is a polynomial of degree $k-1$ in x whose coefficients are determined entirely by the k initial values of (u) and the coefficients of (1.1), *while $F(x)$ is the polynomial associated with* (1.1).

We shall call the polynomial $U(x)$ which completely determines the k initial values of (u) and hence (u) itself, the *generator* of (u).

FUNDAMENTAL THEOREM ON NULL SEQUENCES. *If $U(x)$ is the generator of the sequence (u), then a necessary and sufficient condition that (u) should be a null sequence with numeric less than or equal to n is that*

$$x^n U(x) \equiv 0 \qquad (\mathrm{modd}\ m, F(x)). \tag{5.4}$$

We have the following important corollaries to these theorems.

COROLLARY 1. *If (u) is a purely periodic sequence modulo m, its characteristic number is the least value of n for which the congruence* (5.2) *is satisfied.*

COROLLARY 2. *If (u) is a null sequence modulo m, its numeric is the least value of n for which the congruence* (5.4) *is satisfied.*

The generator of the sequence (w) with the initial values $0, 0, \cdots, 0, 1$ is unity. Hence we have from Theorem 3.1

COROLLARY 3. *The principal period of* (1.1) *modulo m is the least value of n such that*

$$x^n \equiv 1 \qquad (\mathrm{modd}\ m, F(x)).$$

6. We are now ready to establish the isomorphism between the ring of residue classes associated with the double modulus m, $F(x)$ and the group of reduced sequences defined in §4. The ring may be represented by the set of m^k polynomials

$$L(x) = l_0 x^{k-1} + l_1 x^{k-2} + \cdots + l_{k-1} \qquad (0 \leqq l_i < m).$$

On identifying $U(x)$ of (5.3) modulo m with $L(x)$ we obtain the congruences

$$u_r - c_1 u_{r-1} - c_2 u_{r-2} - \cdots - c_r u_0 \equiv l_r \quad (\mathrm{mod}\ m),\ r = 0, \cdots, k-1. \tag{6.1}$$

These congruences have a unique solution

$$u_i \equiv a_i \quad (\mathrm{mod}\ m),\ 0 \leqq a_i < m;\ i = 0, \cdots, k-1.$$

We associate with $L(x)$ the reduced sequence (a) whose initial values are $a_0, \cdots, a_{k-1}$, and write

$$(a) \sim L(x).$$

Since the congruences (6.1) are solvable for the l_r for any m, given (a), we can determine a unique $L(x)$. The correspondence is therefore a reciprocal one.

Suppose that

$$(b) \sim M(x).$$

Then evidently

$$(a + b) \sim L(x) + M(x).$$

If $L(x)\cdot M(x)\equiv N(x)$ (modd m, $F(x)$), we define the reduced sequence (c) associated with $N(x)$ to be the *product* of the sequences (a) and (b). The exact dependence of the elements of (c) upon those of (a) and (b) need not detain us here. If we write $(a)\cdot(b)$ for the product of the sequences (a) and (b), we have then

$$(a)\cdot(b)\sim L(x)\cdot M(x).$$

It is easily verified that the set $\mathfrak{A}$ with the two operations of addition and multiplication just defined satisfies the postulates for a ring*; hence we have the following result:

THEOREM 6.1. *The set $\mathfrak{A}$ of reduced sequences modulo m forms a commutative ring with respect to the operations of addition and multiplication of sequences defined above which is simply isomorphic with the ring $\mathfrak{R}$ of residue classes associated with the double modulus m, $F(x)$.*

If

$$(a): \qquad a_0,\ a_1,\ a_2,\ \cdots$$

is any sequence of $\mathfrak{A}$, the corresponding element of the ring $\mathfrak{R}$ is

$$L(x) = l_0x^{k-1} + l_1x^{k-2} + \cdots + l_{k-1}$$

where

$$l_r \equiv a_r - c_1a_{r-1} - c_2a_{r-2} - \cdots - c_ra_0 \pmod m,\ r = 0, \cdots, k-1.$$

To examine the nature of this correspondence further, we need the following lemma.

LEMMA. *If (u) is a solution of the difference equation* (1.1), *and if $\Delta(u)$ denotes the determinant*

$$\Delta(u) = \begin{vmatrix} u_0, & u_1, & \cdots, & u_{k-1} \\ u_1, & u_2, & \cdots, & u_k \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ u_{k-1}, & u_k, & \cdots, & u_{2k-1} \end{vmatrix}$$

and $U(x)$ the polynomial

$$U(x) = u_0x^{k-1} + (u_1 - c_1u_0)x^{k-2} + (u_2 - c_1u_1 - c_2u_0)x^{k-3} + \cdots$$
$$+ (u_{k-1} - c_1u_{k-2} - \cdots - c_{k-1}u_0),$$

then $(-1)^k\Delta(u)$ is equal to the resultant of $U(x)$ and $F(x)$, where $F(x)$ is the polynomial associated with the difference equation (1.1).

* van der Waerden, *Algebra*, Berlin, 1930, vol. 1, p. 37.

The nature of the proof is sufficiently indicated by the special case $k=3$. The resultant of $U(x)$ and $F(x)$ may then be expressed as the five-rowed eliminant

$$E = \begin{vmatrix} u_0, & u_1 - c_1u_0, & u_2 - c_1u_1 - c_2u_0, & 0, & 0 \\ 0, & u_0, & u_1 - c_1u_0, & u_2 - c_1u_1 - c_2u_0, & 0 \\ 0, & 0, & u_0, & u_1 - c_1u_0, & u_2 - c_1u_1 - c_2u_0 \\ 1, & -c_1, & -c_2, & -c_3, & 0 \\ 0, & 1, & -c_1, & -c_2, & -c_3 \end{vmatrix}.$$

Now perform upon E the operations

$$\text{row } 1 - u_0 \text{ row } 4 - u_1 \text{ row } 5, \quad \text{row } 2 - u_0 \text{ row } 5.$$

The first two elements in the first three rows of E become zero, so that E reduces to the third-order determinant

$$E = -\begin{vmatrix} u_2, & c_2u_1 + c_3u_0, & c_3u_1 \\ u_1, & u_2 - c_1u_1, & c_3u_0 \\ u_0, & u_1 - c_1u_0, & u_2 - c_1u_1 - c_2u_0 \end{vmatrix}.$$

From the difference equation,

$$u_3 = c_1u_2 + c_2u_1 + c_3u_0, \quad u_4 = c_1u_3 + c_2u_2 + c_3u_1.$$

Hence performing upon E successively the operations

$$\text{col } 2 + c_1 \text{ col } 1, \quad \text{col } 3 + c_2 \text{ col } 1 + c_1 \text{ col } 2,$$

we obtain

$$E = -\begin{vmatrix} u_2, & u_3, & c_3u_1 \\ u_1, & u_2, & c_3u_0 \\ u_0, & u_1, & u_2 - c_1u_1 - c_2u_0 \end{vmatrix} = -\begin{vmatrix} u_2, & u_3, & u_4 \\ u_1, & u_2, & u_3 \\ u_0, & u_1, & u_2 \end{vmatrix} = (-1)^3\Delta(u).$$

THEOREM 6.2. *To the units of the ring $\mathfrak{R}$ correspond those sequences of $\mathfrak{A}$ whose characteristic number is the principal period of the difference equation* (1.1) *modulo m, while to the identity element* 1 *of $\mathfrak{R}$ there corresponds the sequence* (w) *with the initial values* $0, 0, \cdots, 0, 1$.

For the units of $\mathfrak{R}$ are represented by those polynomials $L(x)$ such that the resultant of $L(x)$ and $F(x)$ is prime to m. But if $L(x)=U(x)$ is the generator of the sequence (u), we have just seen that $\Delta(u)$ is numerically

equal to the resultant of $L(x)$ and $F(x)$. By the corollary to Theorem 3.1, the characteristic number of all sequences (u) with $\Delta(u)$ prime to m is the same, and equal to the principal period of (1.1) modulo m. The latter part of the theorem follows from the fact that for the sequence (w): $0, 0, \cdots, 0, 1, \cdots$ we have $W(x)=1$.

III. Simplification of the form of the modulus and associated polynomial

7. If $m=p_1^{n_1}\cdots p_r^{n_r}$ is the decomposition of m into its prime factors, then it is easy to see that the ring associated with the double modulus m, $F(x)$ is the direct sum of the r rings associated with the double moduli $p_i^{n_i}$, $F(x)$. We have of course a similar dissection of the ring $\mathfrak{A}$ into a sum of simpler rings. The following important theorem gives the corresponding reduction of the problem of determining the characteristic number and numeric of any sequence modulo m to the case when m is a power of a prime.

Theorem 7.1. *If*

$$m = p_1^{n_1}\cdots p_r^{n_r}$$

is the decomposition of m into its prime factors, then the characteristic number of any sequence modulo m is the least common multiple of its characteristic numbers modulis $p_i^{n_i}$ $(i=1,\cdots,r)$ while its numeric is the maximum of its numerics modulis $p_i^{n_i}$.

It is sufficient to show that if $m=a\cdot b$ where a and b are relatively prime, then the characteristic number of (u) modulo m is the least common multiple of its characteristic numbers modulo a and modulo b, while its numeric modulo m is the greatest of its numerics modulo a and modulo b.

Let

$$(u) \equiv (v) + (w) \qquad (\text{mod } m)$$

be the unique decomposition of (u) into a null sequence (v) and a purely periodic sequence (w). Then since a and b divide m,

$$(u) \equiv (v) + (w) \ (\text{mod } a), \text{ and } (u) \equiv (v) + (w) \ (\text{mod } b).$$

Furthermore (v) is a null sequence modulis a and b and (w) is a purely periodic sequence modulis a and b.

In view of Theorem 4.1, it is sufficient to prove the result for the numeric of (v) and the characteristic number of (w).

Consider first (v), and let $V(x)$ be its generator, ν_m, ν_a and ν_b its numerics modulis m, a and b respectively, and τ the greatest of ν_a and ν_b. Then by the fundamental theorem of §5,

$$x^{\nu_m}V(x) \equiv 0 \text{ (modd } m, F(x)), \quad x^{\nu_a}V(x) \equiv 0 \text{ (modd } a, F(x)),$$
$$x^{\nu_b}V(x) \equiv 0 \text{ (modd } b, F(x)).$$

Thus $x^{\nu_m}V(x) \equiv 0$ (modd $a, F(x)$ (and (modd $b, F(x)$) so that $\nu_m \geqq \tau$. But since a and b are relatively prime,

$$x^{\tau}V(x) \equiv 0 \qquad \text{(modd } ab, F(x))$$

so that $\tau \geqq \nu_m$. Hence $\tau = \nu_m$.

The proof for the characteristic number of (w) is similar and will be left to the reader.*

We shall assume hereafter that $m = p^N$, p a prime, N a given integer.

Now suppose that

$$F(x) \equiv \{\phi_1(x)\}^{t_1} \cdot \{\phi_2(x)\}^{t_2} \cdots \{\phi_s(x)\}^{t_s} \qquad \text{(mod } p)$$

is the unique decomposition of $F(x)$ modulo p into a product of powers of primary irreducible polynomials $\phi(x)$. Then by Schönemann's second theorem† there exists a decomposition of $F(x)$ modulo p^N of the form

$$F(x) \equiv F_1(x) \cdot F_2(x) \cdots F_s(x) \qquad \text{(mod } p^N) \tag{7.1}$$

where

$$F_i(x) \equiv \{\phi_i(x)\}^{t_i} \qquad \text{(mod } p), \ i = 1, 2, \cdots, s,$$

and the polynomials $F_i(x)$ are primary. We shall refer to (7.1) as a Schönemann decomposition of $F(x)$ (modulo p^N).

Corresponding to this decomposition of $F(x)$, we have a decomposition of the ring associated with the double modulus $p^N, F(x)$ into the direct sum of the s rings associated with the moduli $p^N, F_i(x)$. If $U(x)$ is any element of this ring, and

$$U(x) \equiv U^{(i)}(x) \qquad \text{(modd } p^N, F_i(x)), \ i = 1, \cdots, s,$$

where $U^{(i)}(x)$ is of degree less than $F_i(x)$, then $U(x)$ may be uniquely represented as

$$U(x) \equiv B^{(1)}(x)U^{(1)}(x) + B^{(2)}(x)U^{(2)}(x) + \cdots + B^{(s)}(x)U^{(s)}(x) \text{ (modd } p^N, F(x))$$

where the $B^{(i)}(x)$ are of degree less than $F(x)$ and

$$B^{(i)}(x) \equiv 1 \text{ (modd } p^N, F_i(x)),$$
$$\equiv 0 \text{ (modd } p^N, F_j(x)), \quad j \neq i; \ 1 \leqq j \leqq s; \ i = 1, \cdots, s.$$

---

* See Ward, p. 155, Theorem 3.11.

† See Fricke, work cited, §11.

If (u) is the sequence generated by $U(x)$, $(u^{(i)})$ and $(b^{(i)})$ the sequences generated by $U^{(i)}(x)$ and $B^{(i)}(x)$, the analogous decomposition of (u) is

$$(u) \equiv (b^{(1)})\cdot(u^{(1)}) + (b^{(2)})\cdot(u^{(2)}) + \cdots + (b^{(s)})\cdot(u^{(s)}) \pmod{p^N}.$$

The corresponding theorem for the characteristic numbers and numeric of (u) is as follows:

THEOREM 7.2. *Suppose that* (7.1) *is a Schönemann decomposition of* $F(x)$ *modulo* p^N, *and that* $U(x)$ *is a polynomial of degree* $\leq k-1$ *in* x *generating a sequence* (u). *Furthermore suppose that*

$$U(x) \equiv U^{(i)}(x) \qquad (\text{modd } p^N, F_i(x))$$

where $U^{(i)}(x)$ *is a polynomial of degree less than* $F_i(x)$, *and the generator of a sequence* $(u^{(i)})$ *which is a solution of the difference equation whose associated polynomial is* $F_i(x)$.

Then the characteristic number of (u) (modd $p^N, F(x)$) *is the least common multiple of the characteristic numbers of* $(u^{(i)})$ (modd $p^N, F_i(x)$) *and the numeric of* (u) *is the maximum of the numerics of the* $(u^{(i)})$.

Suppose that

$$(u) \equiv (v) + (w) \pmod{p^N} \text{ and } U(x) \equiv V(x) + W(x) \quad (\text{modd } p^N, F(x))$$

are the decompositions of (u) into a null sequence (v) and a purely periodic sequence (w), and the corresponding decomposition of the generator $U(x)$ of (u). Furthermore, suppose that

$$U(x) \equiv U^{(i)}(x),\ V(x) \equiv V^{(i)}(x),\ W(x) \equiv W^{(i)}(x) \quad (\text{modd } p^N, F_i(x))$$

where the polynomials on the right side of the congruences are of lesser degree than $F_i(x)$, and that $(u^{(i)})$, $(v^{(i)})$ and $(w^{(i)})$ are the solutions of the difference equation associated with $F_i(x)$ with the generators $U^{(i)}(x)$, $V^{(i)}(x)$ and $W^{(i)}(x)$ respectively. Then we may write

$$(7.2)\qquad \begin{aligned} (u^{(i)}) &\equiv (v^{(i)}) + (w^{(i)}) && \pmod{p^N}, \\ U^{(i)}(x) &\equiv V^{(i)}(x) + W^{(i)}(x) && (\text{modd } p^N, F_i(x)). \end{aligned}$$

I assert that (7.2) *gives the decomposition of* $(u^{(i)})$ *into its purely periodic and null components*; for if τ and λ are the numeric and characteristic number of (u), we have by the theorems of §§4 and 5

$$x^\tau V(x) \equiv 0, \quad x^\lambda W(x) \equiv W(x) \qquad (\text{modd } p^N, F(x)).$$

Hence

$$(7.3)\qquad x^\tau V^{(i)}(x) \equiv 0,\ x^\lambda W^{(i)}(x) \equiv W^{(i)}(x) \qquad (\text{modd } p^N, F_i(x))$$

so that by the theorems of §5 $(v^{(i)})$ is a null sequence and $(w^{(i)})$ is a purely

periodic sequence. By Theorem 4.1, the numeric of $(v^{(i)})$ and the characteristic number of $(w^{(i)})$ are the numeric and the characteristic number of $(u^{(i)})$. Call this latter number λ_i and let μ be the least common multiple of $\lambda_1, \lambda_2, \cdots, \lambda_s$. From the second congruence in (7.3), $(w^{(i)})$, and hence $(u^{(i)})$, admits the period λ (modd p^N, $F_i(x)$). Hence λ_i divides λ so that μ divides λ. But clearly

$$(x^\mu - 1)W^{(i)}(x) \equiv 0 \qquad (\text{modd } p^N, F_i(x))$$

so that

$$(x^\mu - 1)W(x) \equiv 0 \qquad (\text{modd } p^N, F_i(x)),\ i=1, \cdots, s.$$

Since the resultant of any two distinct $F_i(x)$ is prime to p, these last congruences imply that

$$(x^\mu - 1)W(x) \equiv 0 \qquad (\text{modd } p^N, F(x)).$$

Hence by the fundamental theorem again, λ divides μ so that λ equals μ.

The proof of the result for the numerics is similar and will be omitted here.

8. In the present section, we shall solve completely the problem of determining the null component and the purely periodic component of any sequence (modd p^N, $F(x)$).

Let us assume that the coefficient c_k in (1.1) is divisible by p. Then in the Schönemann decomposition (7.1) one of the $F_i(x)$ must be of the form $x^{t_i}+pV(x)$; let us suppose that it is $F_1(x)$, so that

$$F_1(x) = x^{t_1} + pV(x).$$

The exponent t_1 is simply the number of consecutive coefficients c_k, $c_{k-1}, c_{k-2}, \cdots$ which are divisible by p. Let

$$F'(x) = F_2(x)\cdot F_3(x) \cdots F_s(x),$$

so that Res $\{F_1(x), F'(x)\}$ is prime to p.

By the fundamental theorem of §5, the sequence (u) is a null sequence modulo p^N when and only when the congruence

$$x^n\, U(x) \equiv 0 \qquad (\text{modd } p^N, F(x))$$

is solvable, $U(x)$ denoting as usual the generator of (u). But this congruence is solvable when and only when the two congruences

$$x^n U(x) \equiv 0 \quad (\text{modd } p^N, F_1(x)), \quad x^n U(x) \equiv 0 \quad (\text{modd } p^N, F'(x))$$

are solvable. The first of these congruences is solvable for any $U(x)$, for we may take $n=Nt_1$. The second is solvable when and only when $U(x)\equiv 0$ (modd $p^N, F'(x)$) for Res $\{x, F'(x)\}$ is prime to p. We have thus established the following theorem.

THEOREM 8.1. *If in the Schönemann decomposition modulo p^N of the polynomial $F(x)$ associated with the difference equation* (1.1),

$$F(x) \equiv F_1(x)\cdot F_2(x) \cdots F_s(x) \pmod{p^N}, \tag{7.1}$$

we have $F_1(x)=x^{t_1}+pV(x)$, then a necessary and sufficient condition that a given solution (u) of (1.1) *be a null sequence modulo p^N is that its generator $U(x)$ satisfy the relation*

$$U(x) \equiv 0 \qquad (\text{modd } p^N, F_2(x) \cdots F_s(x)).$$

In this case its numeric is the least value of n such that

$$x^n\, U(x) \equiv 0 \qquad (\text{modd } p^N, F_1(x)). \tag{8.1}$$

We can prove the following result in very much the same manner.

THEOREM 8.2. *With the hypotheses of Theorem* 8.1, *a necessary and sufficient condition that a given solution (u) of* (1.1) *be purely periodic modulo p^N is that its generator $U(x)$ satisfy the relation*

$$U(x) \equiv 0 \qquad (\text{modd } p^N, F_1(x)).$$

The decomposition of (u) into its purely periodic and null components is now easily effected. For since Res $\{F_1(x), F'(x)\}$ is prime to p, we can determine two polynomials $S_1(x)$, $S_2(x)$ such that

$$S_1(x)F_1(x) + S_2(x)F'(x) \equiv U(x) \qquad (\text{modd } p^N, F(x)).$$

Suppose that

$$S_2(x)F'(x) \equiv V(x), \quad S_1(x)F_1(x) \equiv W(x) \qquad (\text{modd } p^N, F(x))$$

where the degrees of $V(x)$ and $W(x)$ do not exceed $k-1$, and let (v) and (w) be the sequences generated by $V(x)$ and $W(x)$ respectively. Then

$$U(x) \equiv V(x) + W(x) \quad (\text{modd } p^N, F(x)), \quad (u) \equiv (v) + (w) \pmod{p^N},$$

and (v) is a null sequence and (w) a purely periodic sequence modulo p^N.

IV. THE DETERMINATION OF THE NUMERIC

9. If (u) is a null sequence modulo p^N, we have just seen that its generator is of the form

$$U(x) \equiv U'(x)\cdot F_2(x) \cdots F_s(x) \pmod{p^N}$$

and that its numeric is the least value of n such that

$$x^n U'(x) \equiv 0 \qquad (\text{modd } p^N, F_1(x)).$$

$F_1(x)$ it will be recalled is of the form $x^{t_1}+pV(x)$. It may happen that $V(x)$ is also divisible by p. To conserve generality, we therefore assume that

$$F_1(x) = x^{t_1} - p^{\sigma_1}\theta(x); \qquad \theta(x) \not\equiv 0 \pmod p; \qquad \theta(x) \text{ of degree less than } t_1.$$

By Schönemann's theorems,* $U'(x)$ has a decomposition modulo p^N of the form

$$U'(x) \equiv p^M G_1(x) U''(x) \pmod{p^N}$$

where

$$M \geqq 0, \qquad G_1(x) = x^{\alpha_1} + p^{\beta_1}\zeta_1(x);$$

$$\text{Res}\,\{G_1(x), U''(x)\} \text{ prime to } p; \qquad \zeta_1(x) \not\equiv 0 \pmod p.$$

It follows immediately that the numeric of (u) is the least value of n such that

$$x^n G_1(x) \equiv 0 \qquad (\text{modd } p^{N-M}, F_1(x)). \tag{9.1}$$

This minimal value may always be calculated in view of the following two theorems:

THEOREM 9.1. *Suppose that a set of polynomials $U(x)$, $G(x)$, $\zeta(x)$ are defined recursively by*

$$\begin{aligned} U_{r-1}(x) &\equiv G_r(x)\overline{U}_{r-1}(x) && (\text{mod } p^{L_{r-1}}),\ r = 1, 2, \cdots, \\ x^{t_1-\alpha_r}G_r(x) &\equiv p^{\rho_r}U_r(x) && (\text{mod } F_1(x)), \\ G_r(x) &= x^{\alpha_r} + p^{\beta_r}\zeta_r(x), \\ L_r &= N - M - (\rho_1 + \rho_2 + \cdots + \rho_r), \end{aligned}$$

where $U_r(x)$ is not divisible by p, $\overline{U}_{r-1}(x)$ is not divisible by x modulo p, and $\zeta_r(x)$ is a polynomial of degree less than α_r not divisible by p, while $U_0(x) = G_1(x)U''(x)$, $\overline{U}_0(x) = U''(x)$. Then the numbers ρ are all positive, and after a finite number of steps, say l, we will either have

$$N \leqq M + \rho_1 + \rho_2 + \cdots + \rho_l \textit{ or } \text{Res}\,\{U_l(x), F_1(x)\} \textit{ prime to } p.$$

Let l now denote the first time one of these alternatives occurs. Then in the first case, the numeric of (u) is $lt_1-(\alpha_1+\alpha_2+\cdots+\alpha_l)$ and in the second case, the numeric is $lt_1-(\alpha_1+\alpha_2+\cdots+\alpha_l)+\nu_l$, where ν_l is the least value of n such that

$$x^n \equiv 0 \qquad (\text{modd } p^{L_l}, F_1(x)). \tag{9.2}$$

* Fricke, work cited, p. 59, p. 65.

THEOREM 9.2. *Suppose that a set of polynomials $\theta(x)$, $\bar{\theta}(x)$ are defined recursively by*

$$\theta_r(x) \equiv (x^{\tau_r} + p^{\gamma_r}\phi_r(x))\bar{\theta}_r(x) \qquad (\operatorname{mod} p^{L_k}),$$

$$x^{t_1-\tau_r}\theta_r(x) \equiv p^{\sigma_r}\theta_{r+1}(x) \qquad (\operatorname{mod} F_1(x)),\ r = 1, 2, \cdots,$$

where $\theta_r(x)$ is not divisible by p, and $\bar{\theta}_r(x)$ is not divisible by x modulo p, $\phi_r(x)$ is a polynomial of degree less than τ_r not divisible by p, while τ_r is the number of consecutive coefficients of the zeroth, first, second, $\cdots$ powers of x in $\theta_r(x)$ which are divisible by p. Then after a finite number of steps, say h, we will either have $L_l \leqq \sigma_1+\sigma_2+ \cdots +\sigma_h$ or $\tau_h=0$ and Res $\{\theta_h(x), F_1(x)\}$ *prime to p.*

Let h denote the first time one of these alternatives occurs. Then in the first case, the least value of n for which the congruence (9.2) *is satisfied is $\bar{\tau}_h = ht_1-(\tau_1+\tau_2 + \cdots +\tau_h)$. In the second case it is $q_h\bar{\tau}_h$, where q_h is the integer next greater than or equal to L_l divided by $\sigma_1+\sigma_2+ \cdots +\sigma_h$.*

The proofs of these theorems are by induction, and are perfectly straightforward though rather lengthy. They will be omitted here, as the important result is that *the numeric may be calculated if we merely know the Schönemann decompositions of $U(x)$ and $F(x)$ quite independently of the calculation of the characteristic number.*

The following results are immediate corollaries of Theorems 9.1 and 9.2.

COROLLARY 1. *If*

$$F(x) \equiv F_1(x) \cdots F_s(x) \qquad (\operatorname{mod} p^N),$$

$$F_1(x) \equiv x^{t_1} - p^{\sigma_1}\theta_1(x) \qquad (\theta_1(x) \not\equiv 0 \operatorname{mod} p)$$

is the Schönemann decomposition of the polynomial $F(x)$ mod p^N *associated with the difference equation* (1.1), *the least upper bound of the numerics of all solutions of* (1.1) *modulo p^N is qt_1, where q is the integer next greater than or equal to N/σ_1.*

COROLLARY 2.* *The least upper bound for the numerics of all difference equations* (1.1) *modulo p^N whose t_1 last coefficients are divisible by p is Nt_1.*

COROLLARY 3. *The least upper bound for the numeric of all difference equations* (1.1) *of order k modulo p^N is Nk.*

V. THE DETERMINATION OF THE CHARACTERISTIC NUMBER

10. In this division of the paper we shall reduce the problem of determining the characteristic number of any solution of (1.1) to its constituents in the sense explained in the introduction. In view of the results of §7, we may

* Due to Engstrom, p. 218, Theorem 9.

assume that $m = p^N$ where p is a prime, and that the associated polynomial $F(x)$ is of the form

$$F(x) = \{\phi(x)\}^s - p\theta(x) \tag{10.1}$$

where it will be recalled that $\phi(x)$ is primary and irreducible modulo p, while $\theta(x)$ is of lesser degree than $F(x)$.

The results of §8 allow us to assume that (u) is purely periodic. Hence by the fundamental theorem of §5, the characteristic number of (u) is the least value of n such that

$$(x^n - 1)U(x) \equiv 0 \qquad (\text{modd } p^N, F(x)), \tag{10.2}$$

where $U(x)$ is the generator of (u).

The following easily established theorem* justifies us in assuming that $U(x)$ is not divisible by p.

Theorem 10.1. *If* (u) *is any solution of the difference equation* (1.1), *the form of* $F(x)$ *being unrestricted, and if the integer* d *is a common factor of the* k *initial values of* (u), *then the characteristic number of* (u) *to any modulus* m *is the characteristic number of* $d^{-1}(u)$ *modulo* (m/l), *where* l *is the greatest common divisor of* m *and* d.

Suppose that λ is the characteristic number of (u) (modd p^N, $F(x)$), so that

$$(x^\lambda - 1)U(x) \equiv 0 \qquad (\text{modd } p^N, F(x)) \tag{10.21}$$

and let p^K be the first elementary divisor of the matrix of the eliminant of $U(x)$ and $F(x)$ corresponding to the prime p. Then I have shown elsewhere† that (10.21) implies that

$$x^\lambda - 1 \equiv 0 \qquad (\text{modd } p^{N-K}, F(x)).$$

Thus λ is a multiple of the principal period of (1.1) modulo p^{N-K}.

Theorem 10.2. *If the first elementary divisor of the matrix of the eliminant of* $U(x)$ *and* $F(x)$ *corresponding to the prime* p *is* p^K, *then the characteristic number of* (u) (modd p^N, $F(x)$), $N > K$, *is a multiple of the principal period of* (1.1) *modulo* p^{N-K}.

This theorem is of some practical importance, as it gives us a lower limit to the characteristic number of any sequence. The extension to composite m and $F(x)$ unrestricted is obvious in view of the results of §7.

* Ward, p. 157, Theorem 5.2.

† These Transactions, vol. 35 (1933), p. 258.

Since $U(x)$ in (10.2) is not congruent to zero modulo p, we may assume that

$$U(x) \equiv \{\phi(x)\}^b\psi(x) \qquad (\text{mod } p),\ a > b \geqq 0,$$

where Res $\{\psi(x), \phi(x)\}$ is prime to p. Then by Schönemann's second theorem,† we have

$$U(x) \equiv U^*(x)V(x) \qquad (\text{mod } p^N)$$

where

$$U^*(x) = \{\phi(x)\}^b + p\xi(x), \quad \xi(x) \text{ of lower degree than } U^*(x), \tag{10.3}$$

and $V(x) \equiv \psi(x)$, mod p.

It follows that *the characteristic number of* (u) *is the least value of* n *such that*

$$(x^n - 1)U^*(x) \equiv 0 \qquad (\text{modd } p^N, F(x)). \tag{10.4}$$

To avoid circumlocutions, we shall refer to this number as the characteristic number of the congruence (10.4).

If $N=1$, we may replace (10.4) by

$$x^n - 1 \equiv 0 \qquad (\text{modd } p, \{\phi(x)\}^{a-b}). \tag{10.5}$$

Suppose that the polynomial $\phi(x)$ is of degree t in x. Then the characteristic number of

$$x^n - 1 \equiv 0 \qquad (\text{modd } p, \phi(x))$$

is a well known quantity in the Galois field theory‡; for it is simply the exponent to which belongs the mark associated with a root of $\phi(x)=0$ in the Galois field of order p^t. We shall regard this number as known to us§; it is a divisor of p^t-1 and hence prime to p and at most equal to p^t-1. Let us denote it by λ. Then there exist polynomials $\phi(x)$ of degree t for which the corresponding λ equals p^t-1; in other words, p^t-1 is not only an upper bound for λ, but it is the least upper bound for λ.

We have then

$$x^\lambda - 1 = \psi(x)\phi(x) + p\zeta(x) \tag{10.6}$$

where $\psi(x)$ and $\zeta(x)$ are polynomials and $\zeta(x)$ is of lower degree than $\phi(x)$. Since the discriminant of $x^\lambda-1$ is prime to p,

$$\psi(x) \not\equiv 0 \qquad (\text{modd } p, \phi(x)). \tag{10.7}$$

---

† Fricke, work cited, pp. 65–66.

‡ See Dickson, *Linear Groups*, Teubner, 1901, Part I.

§ Compare the remarks in §2 of the introduction.

From (10.6),

$$x^{r\lambda} \equiv 1 + r\psi(x)\phi(x) \qquad (\text{modd } p, \phi^2(x)).$$

Hence the characteristic number of

$$x^n \equiv 1 \qquad (\text{modd } p, \phi^2(x))$$

is $p\lambda$. But since

$$x^{p\lambda} \equiv 1 + \{\psi(x)\}^p\{\phi(x)\}^p \qquad (\text{mod } p),$$

$p\lambda$ is also the characteristic number of (10.5) if $2 \leqq a-b \leqq p$.

Proceeding in this manner, we obtain the following result:

THEOREM 10.3. *If $U(x)$ is the generator of a purely periodic solution (u) of the difference equation* (1.1) *whose associated polynomial is of the form*

$$F(x) \equiv \{\phi(x)\}^a \ (\text{mod } p), \ \textit{while} \ U(x) \equiv \{\phi(x)\}^b V(x) \qquad (\text{mod } p),$$

where Res $\{V(x), \phi(x)\}$ *is prime to p and $\phi(x)$ is irreducible modulo p, then the characteristic number of (u) modulo p is $p^q\lambda$ where the integer q is such that*

$$p^{q-1} < a - b \leqq p^q$$

and λ is the least value of n such that

$$x^n \equiv 1 \qquad (\text{modd } p, \phi(x)).$$

THEOREM 10.4. *Under the hypothesis of Theorem* 10.3, *the principal period of* (1.1) *modulo p is $p^r\lambda$ where the integer r is determined by the condition*

$$p^{r-1} < a \leqq p^r$$

and the least upper bound for the principal period is $p^r(p^t-1)$, where t is the degree of the polynomial $\phi(x)$ in x.

We leave the formulation of the corresponding theorems when $F(x)$ is unrestricted in form and m any square-free integer to the reader.

11. We are now in a position to attack (10.4) in the general case when N is greater than one. We have, with the notation of Theorem 10.4,

$$(11.1) \qquad x^{p^r\lambda} - 1 \equiv p^\sigma V(x) \qquad (\text{mod } F(x))$$

where σ is a positive integer, and $V(x)$ is of lesser degree than $F(x)$. If $V(x) = 0$, we shall think of σ as arbitrarily large. If $V(x) \neq 0$, the value of σ is fixed by the condition $V(x) \not\equiv 0 \ (\text{mod } p)$. Then

$$(11.2) \qquad U^*(x)V(x) \equiv p^\rho W(x) \qquad (\text{mod } F(x))$$

where ρ is a positive integer or zero, and $W(x)$ is of lesser degree than $F(x)$. If $W(x) = 0$, we assign an arbitrarily large value to ρ. Otherwise, the value of ρ is fixed by the condition $W(x) \not\equiv 0 \ (\text{mod } p)$.

ρ may be equally well defined as the largest whole number M such that

$$U(x)V(x) \equiv 0 \qquad (\text{modd } p^M, F(x)).$$

Unless $V(x)$ divides $F(x)$ (when $U(x)$ may be taken so that $W(x)=0$), ρ has a definite upper bound† depending only on $V(x)$, $F(x)$ and p.

From (11.1), we deduce that

$$x^{\lambda p^{r+t}} \equiv (1 + p^\sigma V(x))^{p^t} \equiv 1 + p^{\sigma+t}V(x) + \frac{p^{2\sigma+t}(p^t-1)}{1\cdot 2}V^2(x) + \cdots \pmod{F(x)}.$$

Hence from (11.2),

$$U^*(x)(x^{\lambda p^{r+t}} - 1) \equiv p^{\sigma+\rho+t}W(x) + p^{2\sigma+\rho+t}W(x)\frac{(p^t - 1)}{1\cdot 2}V(x) + \cdots \pmod{F(x)},$$

$$U^*(x)(x^{\lambda p^{r+t}} - 1) \equiv p^{\sigma+\rho+t}W(x) \qquad (\text{modd } p^{\rho+\sigma+t+1}, F(x)),$$

save possibly in the case $p=2$, $\sigma=1$, which we shall exclude. From this last congruence, we deduce the following theorems:

THEOREM 11.1. *If p is an odd prime, $N>1$, the characteristic number of the congruence* (10.4) *is $p^r\lambda$ if $N\leqq\rho+\sigma$ and $\lambda p^{r+N-\rho-\sigma}$ if $N\geqq\rho+\sigma$, where ρ and σ are determined by the congruences* (11.1) *and* (11.2).

THEOREM 11.2. *If p is an odd prime, the least upper bound for the characteristic number of the congruence* (10.4) *for all choices of $U^*(x)$ is $p^r\lambda$ if $N\leqq\rho$ and $\lambda p^{r+N-\rho}$ if $N\geqq\rho$, where ρ is determined by the congruence* (11.2).

The fundamental problem of finding the characteristic number of any linear recursive sequence to any modulus m has thus finally reduced to determining the exponents σ and ρ in (11.1) and (11.2). We shall first seek to determine ρ in the case when p is odd and the exponent a in (10.1) is greater than unity.

If u is an indeterminate, and if we let

$$H(u) = u - \frac{u^2}{2} + \cdots - \frac{u^{p-1}}{p-1},$$

$$K(u) = -\frac{u}{2} + \left(1 + \frac{1}{2}\right)\frac{u^2}{3} - \left(1 + \frac{1}{2} + \frac{1}{3}\right)\frac{u^3}{4} + \cdots$$
$$+\left(1 + \frac{1}{2} + \cdots + \frac{1}{p-2}\right)\frac{u^{p-1}}{p-1},$$

$$L(u) = 1 - u + u^2 - \cdots + u^{p-1},$$

$$H^{(r)}(x) = H((\phi\psi)^{p^r}),\ K^{(r)}(x) = K((\phi\psi)^{p^r}),\ L^{(r)}(x) = L((\phi\psi)^{p^r}),$$

† These Transactions, vol. 35 (1933), p. 258.

and, for uniformity of notation,

$$H^{(-1)}(x) = \zeta(x),$$

then it follows by induction on r from (10.6) that for any positive integral value of r,

$$x^{p^r\lambda} \equiv 1 + p\Theta_1(x) + p^2\Theta_2(x) + \{\psi(x)\}^{p^r}\{\phi(x)\}^{p^r} \qquad (\text{mod } p^3),$$

where

(11.3) $\quad \Theta_1(x) = H^{(r-1)}(x),\ \Theta_2(x) = K^{(r-1)}(x) + H^{(r-2)}(x)L^{(r-1)}(x).$

Now by (10.1),

$$\phi^{p^r} = \phi^a\cdot\phi^{p^r-a} = \phi^{p^r-a}(F + p\theta) \equiv p\theta\phi^{p^r-a} \qquad (\text{mod } F(x)).$$

Therefore

(11.4) $$x^{p^r\lambda} \equiv 1 + p(\theta\psi^{p^r}\phi^{p^r-a} + \Theta_1) + p^2\Theta_2 \qquad (\text{modd } p^3, F(x)).$$

On comparing (11.4) and (11.1), we have

(11.41) $$p^{\sigma-1}V(x) \equiv \theta\psi^{p^r}\phi^{p^r-a} + \Theta_1 + p\Theta_2 \qquad (\text{modd } p^2, F(x)).$$

Therefore a necessary and sufficient condition that σ be greater than one is that $\theta\psi^{p^r}\phi^{p^r-a}+\Theta_1\equiv 0$ (modd p, $F(x)$). This congruence is equivalent to

(11.5) $$\theta\psi^{p^r}\phi^{p^r-a} + \psi^{p^{r-1}}\phi^{p^{r-1}} - \tfrac{1}{2}\psi^{2p^{r-1}}\phi^{2p^{r-1}} + \cdots \equiv 0 \quad (\text{modd } p,\ \{\phi(x)\}^a),$$

which may be looked upon as a condition upon $\theta(x)$.

If $p^r-a>p^{r-1}$ or $\theta(x)\equiv 0$ (mod p), the congruence has no solutions. For if it had a solution, we would have

$$\psi^{p^{r-1}} \equiv 0 \qquad (\text{modd } p,\ \phi(x))$$

contradicting (10.7). If $p^r-a\leqq p^{r-1}$ and $\theta(x)\not\equiv 0$ (mod p), (11.5) implies that

$$\theta(x) \equiv 0\ (\text{modd } p,\ \{\phi(x)\}^c),\ \text{where } c = p^{r-1} - p^r + a.$$

If $\theta(x)\equiv 0$ (modd p, $\{\phi(x)\}^{c+1}$), we again obtain a contradiction of (10.7). Hence

$$\theta(x) \equiv \kappa(x)\{\phi(x)\}^c \quad (\text{mod } p), \qquad \kappa(x) \not\equiv 0 \quad (\text{modd } p,\ \phi(x)).$$

On substituting in (11.5), we find that

(11.6) $$\kappa\psi^{p^r-p^{r-1}} + 1 \equiv 0 \qquad (\text{modd } p,\ \{\phi(x)\}^{p^{r-1}}).$$

This criterion can be greatly simplified. For if $y=x^{p^{r-1}}$,

$$\{\psi(x)\}^{p^r-p^{r-1}} \equiv \{\psi(y)\}^{p-1},\ \{\phi(x)\}^{p^{r-1}} \equiv \phi(y) \qquad (\text{mod } p).$$

Hence (11.6) is equivalent to

$$\kappa(x)\{\psi(y)\}^{p-1} + 1 \equiv 0 \qquad (\text{modd } p,\ \phi(y)).$$

Since $\psi(y) \not\equiv 0$ (modd p, $\phi(y)$), there exists a polynomial $\vartheta(y)$ of degree less than $\phi(y)$ such that

$$\vartheta(y)\{\psi(y)\}^{p-1} + 1 \equiv 0 \qquad (\text{modd } p,\ \phi(y)).$$

Hence $\kappa(x) \equiv \vartheta(y)$ (modd p, $\phi(y)$), so that we may take

$$\kappa(x) = \vartheta(x^{p^{r-1}}),$$

where

(11.7) $$\vartheta(x)\{\psi(x)\}^{p-1} + 1 \equiv 0 \qquad (\text{modd } p,\ \phi(x)).$$

If we let

$$\theta_1(x) = \vartheta(x^{p^{r-1}})\{\phi(x)\}^c,\ F_1(x) = \{\phi(x)\}^a - p\theta_1(x),$$

the results we have obtained may be summarized in the following theorem:

THEOREM 11.3. *If p is an odd prime, $a>1$, the exponent σ in* (11.1) *is generally unity. It is always unity if $p^r - a > p^{r-1}$, or if $\theta(x) \equiv 0$* (mod p) *or if $p^r - a > p^{r-1}$, $\theta(x) \not\equiv 0$* (modd p, $\phi(x)$). *It is greater than unity only when $F(x) \equiv F_1(x)$* (mod p^2) *where the polynomial $F_1(x)$ has been defined above.*

The further study of the exceptional case when $F(x) \equiv F_1(x)$ (mod p^2) would take us too far afield and will not be embarked upon here. The theorems of §13 on the determination of ρ when $a=1$ will give the reader an idea of the considerations which apply. We do however gain additional insight into the close relationship between recurring series and higher congruences if we seek to determine the polynomial $\psi(x)$ in (11.7) which must be known (modd p, $\phi(x)$) for $F_1(x)$ to be well defined. It will be recalled that $\psi(x)$ was originally defined as the quotient obtained on dividing $x^\lambda - 1$ by $\phi(x)$. Hence if

$$x^\lambda - 1 \equiv pL(x) \ (\text{modd } p^2,\ \phi^2(x)),\ L(x) \text{ of lesser degree than } \phi^2(x),$$

$\psi(x)$ satisfies the congruence

$$\psi(x) \equiv L(x) \qquad (\text{modd } p,\ \phi(x)).$$

It is sufficient then for our purpose to determine $L(x)$.

Now if we set

$$\phi^2(x) = x^l - d_1x^{l-1} - \cdots - d_l,$$

$$x^n \equiv \sum_{k=1}^{l} w_{n,k}x^{l-k} \qquad (\text{mod } \phi^2(x)),$$

$$w_{n,l+1} = 0 \qquad (n = 0, 1, 2, \cdots),$$

then it is easily verified that the constants $w_{n,k}$ satisfy the following relations:

$$w_{n+1,k} = w_{n,k+1} + d_k w_{n,1} \quad (k = 1, \cdots, l;\ n = 0, 1, 2, \cdots),$$
$$w_{n,k} = \delta_{n,l-k} \qquad (n < l)$$

where $\delta_{n,l-k}$ is the Kronecker δ. It follows without much difficulty that $w_{0,k}, w_{1,k}, w_{2,k}, \cdots$ is a particular solution of the difference equation

$$\Omega_{n+l} = d_1\Omega_{n+l-1} + \cdots + d_l\Omega_n. \tag{11.8}$$

For convenience denote the sequence $w_{0,l-1}, w_{1,l-1}, w_{2,l-1}, \cdots$ whose initial values are $0, 0, \cdots, 0, 1$ simply by (w). Then we may write for a fixed k

$$w_{n,k} = \sum_{j=1}^{l} c_{kj} w_{n+l-j}$$

where the c_{kj} are integers determined by the l equations

$$\sum_{j=1}^{l} c_{kj} w_{n+l-j} = \delta_{n,l-k} \qquad (n = 0, 1, \cdots, l-1).$$

Thus if

$$W_j(x) = \sum_{k=1}^{l} c_{kj} x^{l-k},$$

$W_j(x)$ is a polynomial of degree $l-1$ in x with integral coefficients, which we may regard as known to us. Then

$$x^n = \sum_{k=1}^{l} w_{n,k} x^{l-k} = \sum_{k=1}^{l}\sum_{j=1}^{l} c_{kj} w_{n+l-j} x^{l-k}$$
$$= \sum_{j=1}^{l} w_{n+l-j} W_j(x).$$

Hence

$$pL(x) \equiv w_{\lambda+l-1}W_1(x) + w_{\lambda+l-2}W_2(x) + \cdots + w_\lambda W_l(x) + 1 \quad (\mathrm{mod}\ \phi^2(x))$$

so that $L(x)$ is determined if we know the residues modulo p^2 of the l terms $w_{\lambda+l-1}, w_{\lambda+l-2}, \cdots, w_\lambda$ of the solution $0, 0, \cdots, 0, 1, d_1, \cdots$ of (11.8). There seems to be no way of obtaining these residues short of calculating the whole sequence (w) modulo p^2 step by step out to $\lambda+l$ terms. Such a calculation will at the same time determine λ after at most p^t-1 terms have been found.

12. We are now in a position to study the value of ρ in (11.2) in the general case when $\sigma=1$. We have from (10.3) and (11.41)

$$(12.1)\quad U^*(x)V(x) \equiv \theta\psi^{p^r}\phi^{p^r-a+b} + \phi^b\Theta_1 + p(\xi\psi^{p^r}\phi^{p^r-a} + \xi\Theta_1 + \phi^b\Theta_2) \quad (\text{modd } p^2, F(x)).$$

Hence ρ is greater than zero when and only when

$$\theta\psi^{p^r}\phi^{p^r-a+b} + \phi^b\Theta_1 \equiv 0 \qquad (\text{modd } p, F(x));$$

that is, when and only when

$$(12.2)\quad \theta\psi^{p^r-a+b} + \phi^{p^{r-1}+b}\psi^{p^{r-1}}(1 - \tfrac{1}{2}\phi^{p^{r-1}}\psi^{p^{r-1}} + \cdots) \equiv 0 \quad (\text{modd } p, \{\phi(x)\}^a).$$

If $p^r-a+b\geqq a$, $p^{r-1}+b\geqq a$, (12.2) is satisfied for any choice of $\theta(x)$. In the contrary case, it is either insolvable or imposes a condition upon $\theta(x)$. We find in fact that there are no solutions in any one of the five following cases:

(i) $\quad p^r - a + b \geqq a,\ p^{r-1} + b < a;$

(ii) $\quad p^r - a + b < a,\ p^{r-1} + b < a,\ p^r - a > p^{r-1};$

(iii) $\quad \theta(x) \equiv 0 \pmod p,\ p^{r-1} + b < a;$

(iv) $\quad p^r - a + b < a,\ p^{r-1} + b < a,\ p^{r-1} \geqq p^r - a,$

$$\theta(x) \not\equiv \kappa(x)\{\phi(x)\}^{p^{r-1}-p^r+a} \qquad (\text{mod } p),$$

where $\kappa(x)\{\psi(x)\}^{p^r-p^{r-1}}+1\equiv 0$ (modd p, $\{\phi(x)\}^{2a-p^r-b}$);

(v) $\quad p^r - a + b < a,\ p^{r-1} + b \geqq a,\ \theta(x) \not\equiv 0 \quad (\text{modd } p, \{\phi(x)\}^{2a-p^r-b}).$

Thus generally speaking, if $\sigma=1$, $\rho=0$ unless $b\geqq a-p^{r-1}$, $b\geqq 2a-p^r$. Passing to this case, we have from (10.1), (11.21) and (12.1)

$$\begin{aligned} U^*(x)V(x) \equiv p\{&\theta^2\psi^{p^r}\phi^d + \theta\psi^{p^{r-1}}\phi^e(1 - \tfrac{1}{2}\psi^{p^{r-1}}\phi^{p^{r-1}} + \cdots) + \xi\psi^{p^r}\phi^{p^r-a} \\ &+ \xi\psi^{p^{r-1}}\phi^{p^{r-1}}(1 - \tfrac{1}{2}\psi^{p^{r-1}}\phi^{p^{r-1}} + \cdots) \\ &+ \phi^{b+p^{r-1}}\psi^{p^{r-1}}(-\tfrac{1}{2} + \tfrac{1}{2}\phi^{p^{r-1}}\psi^{p^{r-1}} - \cdots) \\ &+ \phi^{b+p^{r-2}}\psi^{p^{r-2}}(1 - \tfrac{1}{2}\psi^{p^{r-2}}\phi^{p^{r-2}} + \cdots)\} \quad (\text{modd } p^2, F(x)), \end{aligned}$$

where the last group of terms within the bracket must be replaced by $\phi^b\zeta(x)(1-\psi\phi+\cdots)$ if $r=1$, and the exponents d and e in the first two groups of terms are $\geqq 0$ and have the values p^r-2a+b, $p^{r-1}+b-a$.

Hence $\rho=1$ unless the expression in brackets above is congruent to zero (modd p, $F(x)$) or

$$(12.3)\quad \begin{aligned} &\theta^2\psi^{p^r}\phi^d + \theta\psi^{p^{r-1}}\phi^e + \xi\psi^{p^r}\phi^{p^r-a} + \xi\psi^{p^{r-1}}\phi^{p^{r-1}} + \phi^{b+p^{r-1}}\psi^{p^{r-1}} + \phi^{b+p^{r-2}}\psi^{p^{r-2}} \\ &+ \phi^{b+2p^{r-2}}E + \xi\phi^{2p^{r-1}}F + \phi^{b+2p^{r-1}}G + \theta\phi^{e+p^{r-1}}H \equiv 0 \quad (\text{modd } p, \phi^a), \end{aligned}$$

where E, F, G, H denote polynomials in x which are not congruent to zero (modd p, $\phi(x)$) with integral coefficients modulo p. The terms $\phi^{b+2p^{r-2}}E + \phi^{b+p^{r-2}}\psi^{p^{r-2}}$ must be replaced by $\phi^b\zeta + \phi^{b+1}\zeta E$ if $r=1$.

It is not difficult to show that the lowest exponent of ϕ occurring in (12.3) is either d or e so that (12.3) imposes a condition upon $\theta(x)$ of the type appearing under (12.2),

$$\theta(x) \equiv \{\phi(x)\}^g \kappa(x) \qquad (\text{mod } p).$$

The exponent g here depends upon the relative magnitudes of a, b, p^r, p^{r-1}, p^{r-2} but may be shown to be positive. We may therefore state the following theorem:

THEOREM 12.1. *If p is an odd prime, $F(x) = \{\phi(x)\}^a + p\theta(x)$, $a>1$, $\theta(x) \not\equiv 0$ (modd p, $\phi(x)$), then ρ in (11.2) is unity if $p^r + b \geqq 2a$, $p^{r-1}+b \geqq a$ and zero otherwise. If $\theta(x) \equiv 0$ (mod p), ρ is zero if $p^{r-1}+b<a$, and if $p^{r-1}+b \geqq a$ it is unity unless both $p^r - a$ and p^{r-1} are $\leqq b+p^{r-2}$ and $\theta(x)$ satisfies a special condition. If $\theta(x) \equiv 0$ (modd p, $\phi(x)$) $\not\equiv 0$ (mod p), the same results usually apply unless $F(x)$ is of a special form similar to that of $F_1(x)$ in Theorem* 11.4.

13. We shall conclude by discussing the case when the exponent a in (10.1) is unity so that

$$F(x) = \phi(x) - p\theta(x). \tag{13.1}$$

A necessary condition for this to hold is that p should not divide the discriminant of $F(x)$. Hence if this discriminant is not zero, the results of this section will apply to the powers of all primes save a finite number.

If the sequence (u) is not divisible by p, Res $\{U(x), F(x)\}$ is necessarily prime to p, so that the characteristic number of (u) modulo p^N is the principal period of (1.1), and hence the characteristic number of the congruence

$$x^n \equiv 1 \qquad (\text{modd } p^N, F(x)).$$

With the notation of §10, let λ be the characteristic number of the congruence

$$x^n \equiv 1 \qquad (\text{modd } p, \phi(x)),$$

so that we have identically in x

$$\begin{aligned} x^\lambda - 1 &= \psi(x)\phi(x) + p\zeta(x), \\ \psi(x) &\not\equiv 0 \end{aligned} \qquad (\text{modd } p, \phi(x)). \tag{13.2}$$

We shall now establish the following comprehensive theorem:

THEOREM 13.1. *Let p be an odd prime, $\phi(x)$ an irreducible polynomial modulo p, and suppose that the polynomial $F(x)$ associated with the difference equation* (1.1) *is of the form* (13.1). *Furthermore, let*

$$F_2(x) = \phi(x) - p\theta_1(x)$$

where $\xi(x) = \theta_1(x)$ is a solution of the congruence

$$\psi(x)\xi(x) + \zeta(x) \equiv 0 \qquad (\text{modd } p, \phi(x)),$$

$\psi(x)$ and $\zeta(x)$ being given† *by* (13.2) .

Then if $F(x) \not\equiv F_2(x)$ (mod p^2), *the characteristic number modulo p^N of any solution of* (1.1) *which is not divisible by p is $p^{N-1}\lambda$, where λ is the least value of n such that*

$$x^n \equiv 1 \qquad (\text{modd } p, \phi(x)).$$

On the other hand, if $F(x) \equiv F_2(x)$, mod p^2, there exists a set of polynomials $F_2(x), F_3(x), \cdots, F_T(x), \cdots$, depending only upon p, $\phi(x)$, $\psi(x)$ and $\zeta(x)$, such that if $F(x) \equiv F_T(x)$ (mod p^T), *$\not\equiv F_{T+1}(x)$* (mod p^{T+1}), *the characteristic number is λ or $p^{N-T}\lambda$ according as $N \leqq T$ or $N \geqq T$.*

We have

$$x^\lambda - 1 = \psi(x)F(x) + p(\theta(x)\psi(x) + \zeta(x)).$$

Suppose first that $\theta(x)\psi(x) + \zeta(x) \not\equiv 0$ (modd p, $\phi(x)$). Then

$$x^\lambda \equiv 1 + pK(x) \qquad (\text{mod } F(x))$$

where $K(x)$ is of lesser degree than $F(x)$ and not divisible by p. On raising this last congruence to the p^rth power, we obtain

$$x^{p^r\lambda} \equiv 1 + p^{r+1}K(x) + \frac{p^r(p^r-1)}{1\cdot 2}p^2K(x) + \cdots \qquad (\text{mod } F(x)). \tag{13.3}$$

Hence if p is an odd prime,

$$x^{p^r\lambda} \equiv 1 + p^{r+1}K(x) \qquad (\text{modd } p^{r+2}, F(x)).$$

But clearly

$$x^{p^{r+1}\lambda} \equiv 1 \qquad (\text{modd } p^{r+2}, F(x)).$$

Since the characteristic number of (13.1) for $N = r+2$ is a multiple of its characteristic number for $N = r+1$, it is exactly equal to $p^{N-1}\lambda$.

Now let us assume that

---

† They may be determined sufficiently to define $F_2(x)$ by the procedure sketched in §11.

$$\psi(x)\theta(x) + \zeta(x) \equiv 0 \qquad (\text{modd } p,\ \phi(x)).$$

This congruence has a unique solution modulo p of degree less than $\phi(x)$. Let us denote it by $\theta_1(x)$, and set

$$F_2(x) = \phi(x) - p\theta_1(x).$$

Then if $F(x) \not\equiv F_2(x)$ (mod p^2), $\theta(x) \not\equiv \theta_1(x)$ (mod p). Consequently $\psi(x)\theta(x) + \zeta(x) \not\equiv 0$ (modd p, $\phi(x)$) and the argument just given is applicable. Assume then that

$$F(x) \equiv F_2(x) \qquad (\text{mod } p^2).$$

Consider the polynomials

$$F_1(x), F_2(x), F_3(x), \cdots, F_k(x), \cdots$$

defined by the recursive relations†

$$F_k(x) = \phi(x) - p\Theta_{k-1}(x), \quad \Theta_k(x) = \Theta_{k-1}(x) + p^{k-1}\theta_k(x), \quad \Theta_0(x) = 0,$$

$$(13.4) \qquad \begin{aligned} &\psi(x)\Theta_{k-1}(x) + \zeta(x) \equiv p^{k-1}r_k(x) && (\text{modd } p^k, F_k(x)), \\ &\psi(x)\theta_k(x) + r_k(x) \equiv 0 && (\text{modd } p, \phi(x)),\ k = 1, 2, 3, \cdots. \end{aligned}$$

These relations are consistent with one another; for if $k=1$ they give $F_1(x) = \phi(x)$ and for $k=2$ they give the polynomial $F_2(x)$ defined above. If we assume that they are consistent for $k=1, 2, 3, \cdots, s$ it easily follows that they are consistent for $k=s+1$.

Now suppose that

$$F(x) \equiv F_T(x) \ (\text{mod } p^T), \not\equiv F_{T+1}(x) \ (\text{mod } p^{T+1}),\ T \geqq 2.$$

Then

$$x^\lambda - 1 \equiv 0 \ (\text{modd } p^T, F(x)), \not\equiv 0 \ (\text{modd } p^{T+1}, F(x)).$$

For by (13.2) and the relations (13.4),

$$\begin{aligned} x^\lambda - 1 &= \psi(x)\phi(x) + p\zeta(x) = \psi(x)\{F_T(x) + p\Theta_{T-1}(x)\} + p\zeta(x) \\ &= \psi(x)F_T(x) + p(\psi(x)\Theta_{T-1}(x) + \zeta(x)) \\ &\equiv p(\psi(x)\Theta_{T-1}(x) + \zeta(x)) && (\text{mod } F_T(x)) \\ &\equiv p\cdot p^{T-1}r_{T-1}(x) && (\text{modd } p^T, F_T(x)) \\ &\equiv 0 \quad (\text{modd } p^T, F_T(x)), \quad \equiv 0 && (\text{modd } p^T, F(x)). \end{aligned}$$

In like manner it can be shown that

$$x^\lambda - 1 \not\equiv 0 \qquad (\text{modd } p^{T+1}, F(x)).$$

† The $\Theta(x)$ here have no connection with those of §11.

Hence we have

$$x^{\lambda} \equiv 1 + p^{T}K(x) \qquad (\mathrm{mod}\, F(x)),$$

where $K(x) \not\equiv 0$ (modd p, $F(x)$). On raising this congruence to the appropriate power, we find that whether p be even or odd the characteristic number is $p^{N-T}\lambda$ or λ according as $N \geqq T$ or $N \leqq T$.

The case $p=2$, $T=1$ demands separate treatment. If $\theta(x)\psi(x)+\zeta(x) \equiv K(x) \not\equiv 0$ (modd 2, $\phi(x)$), we obtain from (12.3), on putting $p=2$,

$$x^{2^{r}\lambda} = 1 + 2^{r+1}K(x)(1 + (2^{r} - 1)K(x) + \cdots) \equiv 1 + 2^{r+1}K(x)(1 - K(x)) \qquad (\mathrm{modd}\ 2^{r+2}, F(x)).$$

If $K(x) \not\equiv 1$ (mod 2), the previous argument for p odd is applicable. But in case $K(x) \equiv 1$ (mod 2), the characteristic number is a divisor of $2^{r}\lambda$.

Since $K(x)$ is of lesser degree than $F(x)$, the most general assumption is that

$$K(x) + 1 = 2^{s}L(x) \text{ where } L(x) \not\equiv 0 \qquad (\mathrm{mod}\ 2).$$

Then

$$x^{\lambda} \equiv -1 + 2^{s+1}L(x) \qquad (\mathrm{mod}\, F(x)), \tag{13.5}$$

$$x^{2\lambda} \equiv 1 \qquad (\mathrm{modd}\ 2^{s+2}, F(x)).$$

Hence if $N=1$, the characteristic number is λ, while if $s+2 \geqq N > 1$, the characteristic number is 2λ. On raising (13.5) to a power of 2, we find that if $N \geqq s+2$, the characteristic number is $2^{N-s-1}\lambda$.

These results determine the characteristic number in the excluded case of (11.1) when $\sigma=1$ and $p=2$ for all $F(x)$ of the form $\phi(x)-2\theta(x)$. The further discussion of the characteristic number for powers of 2 demands a special treatment which will be given elsewhere.

California Institute of Technology,
Pasadena, Calif.

SETS OF k-EXTENT IN n-DIMENSIONAL SPACE†

BY

R. L. JEFFERY

1. Introduction. Let A be any point set on the bounded n-dimensional domain D. In his development of the theory of measure C. Carathéodory‡ has defined in connection with the set A a measurable set $\overline{A}$ which has come to be called the *massgleiche Hülle* of A.§

Let B_n be a sequence of open sets containing A such that $B_n \supset B_{n+1}$, and $\lim \mu B_n = \mu^* A$. Then the set

$$\overline{A} = B_1 B_2 \cdots$$

contains A, and is measurable with

$$\mu \overline{A} = \mu^* A.$$

The number of ways of selecting each B_n is more than countable, and no rule is given for any particular choice. Consequently it is impossible to say of every point of the domain D whether or not it belongs to the set $\overline{A}$. This amounts to saying that the set $\overline{A}$ is not well-defined.

It is possible to replace the set A by a set A' which is effectively defined. Let B be the complement of A on D. Let ω_k be a sequence of cells with a point b of B as center, and with equal side lengths tending to zero as k increases. Let

$$\rho(b, \omega_k) = \frac{\mu^* A \omega_k}{\mu^* B \omega_k}\,.$$

Since at each point of B except at most a null set the outer metric density of B is unity, it follows that $\rho(b, \omega_k)$ is defined for all values of k at almost all points of B. Let C be the part of B for which $\rho(b, \omega_k)$ is defined for all values of k, and for which

$$\overline{\lim_{k\to\infty}}\, \rho(b, \omega_k) > 0.$$

The set

$$A' = A + C$$

† Presented to the Society, March 25, 1932; received by the editors June 14, 1932, and, in revised form, November 18, 1932.

‡ *Über das lineare Mass von Punktmengen*, Göttinger Nachrichten, 1914.

§ Hahn, *Theorie der Reellen Funktionen*, p. 435. Carathéodory, *Vorlesungen über reelle Funktionen*, p. 260.

contains A, and is effectively defined in terms of A. We also have the following:

I. The set A' is measurable in the sense of Carathéodory, and

$$\mu A' = \mu^* A.$$

II. A necessary and sufficient condition that A be measurable is that

$$\mu^* C = 0.$$

Though not explicitly stated, the proofs of I and II are contained in a previous discussion.†

An analogous situation exists in connection with plane sets of linear extent. Let A be such a set with finite outer linear measure‡ equal to l. Let $U_n = u_{n1}, u_{n2}, \cdots$ be a sequence of open convex areas which contains A, with d_{ni}, the greatest diameter of u_{ni}, tending to zero, and with $\sum d_{ni}$ tending to l. Then the set

$$\overline{A} = U_1 U_2 \cdots$$

contains A, is linearly measurable,§ and

$$L\overline{A} = L^* A.$$

In this case too there is no way to determine for each point of the domain D containing A whether or not it belongs to the set $\overline{A}$. In the present paper we determine a set A' which contains the plane set of linear extent A, which is linearly measurable, with

$$LA' = L^* A,$$

and which is well-defined in terms of A.

That these sets A' are well-defined has some significance.‖ A more important consideration is, however, that the concepts involved in and leading up to their definition combine to form an elegant and very useful tool for handling certain types of problems.¶ We do not restrict ourselves to plane sets, but carry through the discussion for sets of extent k in n-dimensional space. We show that such sets have properties of density similar to the properties of density which Besicovitch†† and Sierpinski‡‡ respectively have shown to hold

† Annals of Mathematics, (2), vol. 33, pp. 449–451.

‡ Carathéodory, Göttinger Nachrichten, loc. cit., §23.

§ Carathéodory, Göttinger Nachrichten, loc. cit., §28.

‖ Sierpinski, Fundamenta Mathematicae, vol. 2, pp. 112.

¶ See Annals of Mathematics, (2), vol. 33, pp. 452–459, these Transactions, vol. 34, p. 650, also the concluding section of this paper.

†† Mathematische Annalen, vol. 98, p. 422.

‡‡ Fundamenta Mathematicae, vol. 9, p. 172.

for linearly measurable plane sets, and for plane sets which are not necessarily linearly measurable but which have linear extent. Although there would be no difficulty in giving independent proofs of the various results, to conserve space we have, whenever possible, based our proofs on those of Carathéodory and Besicovitch.

Let S_n be the n-dimensional euclidean space, and S_k a k-dimensional flat space† in S_n. Let U be an open convex† domain of S_n. For a given U let S_k be such that the k-dimensional measure of S_kU is a maximum, and denote this maximum measure by l_k. We shall call l_1 the greatest diameter of U, and denote it by d. Let A be any bounded set in S_n, and ρ any positive number. Put A in a countable set of open convex domains u_i. Let $L_k^\rho(A)$ be the lower bound of $\sum l_k^i$ for all possible such enclosures with $d_i<\rho$. Evidently $L_k^\rho(A)$ does not decrease as ρ decreases. Let

$$L_k(A) = \lim_{\rho\to 0} L_k^\rho(A).$$

It is clear that $L_k(A)\geqq 0$, and may be infinite.

The largest value of k for which $L_k(A)\neq 0$ determines the extent of the set A, and the number $L_k(A)$, finite or infinite, is the outer k-dimensional measure of the set A. If for each arbitrary set W of extent k

$$L_k(W) = L_k(AW) + L_k(W - AW),$$

the set A is measurable. This definition of measurability, which is based on that of Carathéodory for sets of linear extent, coincides with that of Lebesgue for n-dimensional sets. But not all such sets are measurable in the sense of Lebesgue. Likewise not all sets of extent k are measurable in the sense of Carathéodory. An obvious example is a linear set in the plane which is non-measurable in the sense of Lebesgue.

The theory developed by Carathéodory for linear outer measure, and for measurability when the set A is measurable, is easily shown to hold for the measure function $L_k(A)$. For convenient reference we recall such results of this theory as we shall have occasion to use.

CI. *If the sets A and B are of extent k, and if A contains B, then*

$$L_k(B) \leqq L_k(A).$$

CII. *If A is the set each point of which is on one of the sets $A_1, A_2, \cdots$, then*

$$L_k(A) \leqq L_k(A_1) + L_k(A_2) + \cdots.$$

† A space which by a proper choice of coordinate axes can be represented by $x_1=x_2=\cdots=x_{n-k}=0$. A domain U is convex if every S_2U is convex.

CIII. *If A_1, A_2, $\cdots$ is a sequence of sets such that A_n contains A_{n-1}, and A is the limit set, then*

$$\lim_{n\to\infty} L_k(A_n) = L_k(A).$$

CIV. *If A and B are such that every point of A is a distance not less than $\delta>0$ from any point of B, then*

$$L_k(A) + L_k(B) = L_k(A + B).$$

2. Some general lemmas. In this section we prove three lemmas.

LEMMA I. *If A is such that $L_k(A)$ is finite and different from zero, then $L_{k-1}(A)$ is infinite and $L_{k+1}(A)=0$.*

That $L_{k+1}(A)=0$ follows readily from the fact that, for any set of domains u_i with $d_i<\rho$, $\sum l_{k+1}^i<\rho\sum l_k^i$. We then have $L_{k-1}(A)$ infinite. For a supposition that $L_{k-1}(A)$ is finite makes $L_k(A)=0$.

LEMMA II.† *Let $V=V_1$, V_2, $\cdots$ be an infinite sequence of open convex domains in S_n, and A any set of points. Then*

$$L_k(AV) + L_k(A - AV) = L_k(A).$$

First let V consist of a single domain, and let U_1, U_2, $\cdots$ be a sequence of closed domains interior to V and such that U_n contains U_{n-1} and $\lim U_n=V$. Let $A_n=AU_n$. Then $\lim A_n=AV$. The sets A_n and $A-AV$ are on closed mutually exclusive domains. Hence these two sets satisfy the conditions of CIV, and it follows that

$$L_k(A_n) + L_k(A - AV) = L_k(A_n + A - AV) \leqq L_k(A).$$

And since by CIII $\lim L_k(A_n)=L_k(AV)$ we have

$$L_k(AV) + L_k(A - AV) \leqq L_k(A).$$

But by CII

$$L_k(AV) + L_k(A - AV) \geqq L_k(A).$$

These two inequalities give the Lemma for V a single region. The extension to the case where V consists of a finite number of regions is obvious. When $V=v_1, v_2, \cdots$, set $V_n=v_1, v_2, \cdots, v_n$. Then

$$L_k(AV_n) + L_k(A - AV_n) = L_k(A). \tag{1}$$

† It has been remarked by Mr. J. F. Randolph that Lemma II follows from the definition of Carathéodory for k-dimensional measurability, provided the open set V in S_n is considered to be k-dimensional measurable in the sense of Carathéodory, with infinite measure if $k<n$. In this connection we note that if $k<n$ every open set V in S_n does not satisfy the criterion of measurability which is obtained for sets of finite extent in Theorem XII of the present paper.

The set $A-AV_n$ tends to $A-AV$, and the set AV_n tends to AV. And since AV_n contains AV_{n-1} it follows from CIII that

$$\lim L_k(AV_n) = L_k(AV).$$

Hence

$$L_k(AV) + \lim L_k(A - AV_n) = L_k(A). \tag{2}$$

It follows from CI that

$$L_k(A - AV) \leqq \lim L_k(A - AV_n).$$

Suppose the equality sign does not hold. Then from (2) we get

$$L_k(AV) + L_k(A - AV) < L_k(A),$$

which, by CII, is not true. Hence

$$L_k(AV) + L_k(A - AV) = L_k(A),$$

and the Lemma is proved.

LEMMA III. *Let $V(\rho)$ denote any finite or countably infinite set of open convex domains $V_1, V_2, \cdots$ with greatest diameter $d_i<\rho$. Then to any set A of extent k and any positive number ϵ there corresponds a number $\rho_1>0$ such that for any set $V(\rho)$ with $\rho<\rho_1$ the inequality*

$$L_k[AV(\rho)] < \sum l_k^i + \epsilon$$

is satisfied.

This Lemma has been proved for linearly measurable plane sets by Besicovitch.† His inequality (2) follows from the measurability of the set. But the corresponding inequality for any set follows from Lemma II above. The remainder of the argument is similar to that of Besicovitch with $\sum l_k^i$ replacing $\sum d_i$ for the various regions involved.

3. **Density.** Let A be a set of extent k, a any point of A, and $H(a, r)$ an n-dimensional hypersphere with center a and radius r. Let h_k^r be the k-dimensional measure of the maximal k-dimensional flat space that can be inscribed in $H(a, r)$. Let

$$D(a, r) = \frac{L_k[AH(a, r)]}{h_k^r},$$

and let $D^*(a)$ and $D_*(a)$ be the upper and lower limits respectively of $D(a, r)$ as r tends to zero. These numbers are respectively the upper and lower densities of A at a.

† Loc. cit., p. 427.

It has been shown by Besicovitch† that the linear measure of a plane set depends on the type of region u_i used in estimating this measure. A similar state of affairs is to be expected for sets with extent greater than unity. In estimating outer measure we shall take into consideration only types of regions u_i which are such that

$$l_k^i = \phi_k d^k,$$

where ϕ_k is a constant depending on k and on the particular type of region, and d is the greatest diameter of the region. For a hypersphere $H(a, r)$ let

$$h_k^r = \psi_k (2r)^k.$$

We then have

THEOREM I. *If A is any set of extent k, then for almost all points of A,*

$$\frac{1}{2^k}\frac{\phi_k}{\psi_k} \leqq D^*(a) \leqq 1.$$

Let E be the part of A for each point of which $D^*(a)>1$. Let E_λ be the part of E about each point e of which there exists a sequence of hyperspheres $H(e, r_i)$ with

$$\frac{L_k[AH(e, r_i)]}{h_k^{r_i}} > 1 + \lambda. \tag{1}$$

By Lemma III there exists $\rho>0$ such that for any set of hyperspheres H_i with $r_i<\rho$ we have

$$\sum L_k(E_\lambda H_i) < \sum h_k^{r_i} + \epsilon. \tag{2}$$

From the set of hyperspheres defined in (1) let those with $r_i>\rho$ be discarded. It is then possible to use Vitali's argument to show the existence of a countable non-overlapping sequence of the remaining hyperspheres of H_i which contain almost all of E_λ. From (1) for this sequence we have

$$\sum h_k^{r_i}(1 + \lambda) < L_k(E_\lambda). \tag{3}$$

But from (2) we get

$$\sum h_k^{r_i} > L_k(E_\lambda) - \epsilon. \tag{4}$$

But for λ sufficiently small $L_k(E_\lambda)>0$, and ϵ can be taken arbitrarily small independently of λ. This makes (3) and (4) contradictory, which proves that $D^*(a)\leqq 1$ for almost all A.

† Loc. cit., p. 459.

To complete the proof of Theorem I we notice that the argument of Besicovitch† may be used to establish the existence of a part A_1 of A with $L_k(A_1)$ arbitrarily near to $L_k(A)$ such that about each point a_1 of A_1 there exists a hypersphere $H(a_1, d)$ for which

$$L_k\{AH(a_1, d)\} \geqq \frac{\phi_k d^k}{1+\eta},$$

where η is arbitrary and $d<\eta$.

From this we get, by dividing by $h_k{}^d$,

$$\frac{L_k[AH(a_1, d)]}{h_k{}^d} \geqq \frac{\phi_k d^k}{h_k{}^d} - \eta \geqq \frac{1}{2^k}\frac{\phi_k}{\psi_k} - \eta\,.$$

Completing the argument along the lines followed by Besicovitch we finally arrive at

$$D^*(a) \geqq \frac{1}{2^k}\frac{\phi_k}{\psi_k}$$

at almost all points of A.

We note that if hyperspheres are used in computing the outer measure of A then we get $1/2^k$ as a lower bound for the upper density at almost all points of A.

If, at a point a of the set A, $D_*(a) = D^*(a) = 1$, then A is *regular* at a. Otherwise A is *irregular* at a. The existence of sets of extent k which are regular at almost every point is obvious. Besicovitch‡ has shown the existence of linearly measurable plane sets which are irregular. An evident modification of his methods may be used to construct sets of extent $k=2$ in S_3 which are irregular.

THEOREM II. *If A is any set of extent k then the part of A for which $D^*(a)=0$ has zero measure, regardless of the type of region used in estimating $L_k(A)$.*

For the sake of simplicity we prove this for a set A of extent two in three-dimensional space. With suitable notation the method may be used to obtain the same result for sets of extent greater than two.

Let A_δ be the part of A for which

$$\frac{L_2[AH(a, r)]}{h_2{}^r} < \epsilon,\ r < \delta.$$

For δ sufficiently small $L_2(A_\delta) > 0$. Put A_δ in $u_1, u_2, \cdots$ where $d_i < \delta/2^{3/2}$, and $\sum l_2^i < L_2(A_\delta) + \epsilon$. About u_i circumscribe a rectangular parallelepiped p_i with

† Loc. cit., pp. 428–429.

‡ Loc. cit., p. 431.

longest side parallel to a greatest diameter d_i of u_i. Then, since u_i is convex, the maximal plane section of p_i is not greater than $4l_2^i$. Circumscribe p_i by a cylinder C_i with axis parallel to longest side of p_i. Then the measure q_i of a cross section of this cylinder through the axis is not greater than $4l_2$. It is evidently possible to cover a part of C_i with cylinders C_{ij} with length t_{ij} and radius r_{ij} where $t_{ij}=2r_{ij}\leqq 2d_i$, where $\sum q_{ij}\leqq 2q_i\leqq 8l_2$, and where $\sum_i\sum_j L_k(A_\delta C_{ij})\geqq L_k(A_\delta)/2$. Fix any point of A_δ in C_{ij}, and with this point as center construct a sphere $H(a_\delta, 2t_{ij})$ with radius $2t_{ij}$. $H(a_\delta, 2t_{ij})$ then contains C_{ij} and $h_2^{2t_{ij}}=4\pi(2t_{ij})^2=16\pi q_{ij}$. From these and (1), since the greatest linear dimension of C_{ij} is less than $2^{3/2}d_i<\delta$, we get

$$\frac{L_2[A_\delta C_{ij}]}{h_2^{2t_{ij}}}<\epsilon,$$

which gives

$$\sum_i\sum_j L_2(A_\delta C_{ij})\leqq \epsilon\sum_i\sum_j h_2^{2t_{ij}}\leqq \epsilon 16\pi\sum_i\sum_j q_{ij}\leqq \epsilon 128\pi\sum_i l_2^i,$$

$$\frac{\lambda}{2}<\epsilon 128\pi[L_k(A_\delta)+\epsilon]<\frac{\lambda+\epsilon}{4}\cdot$$

But λ can be fixed greater than zero, and ϵ can be taken arbitrarily small independent of λ. We are thus led to a contradiction, which proves the theorem.

4. **Separated sets.** A point set A is separated from a point set B, if a part of A can be put in a set of open convex regions α in such a way that

$$\text{(1)}\qquad L_k(\alpha B)<\epsilon,\ \text{and}\ L_k(A-\alpha A)<\epsilon.$$

THEOREM III. *If A is separated from B, then B is separated from A.*

It follows from CIII that if $\alpha=\alpha_1, \alpha_2, \cdots$ satisfies (1), and if $\alpha'=\alpha_1, \alpha_2, \cdots, \alpha_n$, then for n sufficiently large

$$L_k(\alpha' A)>L_k(A)-2\epsilon.$$

Let $V_i=v_{i1}, v_{i2}, \cdots, v_{in}$ where v_{ij} is a closed region interior to α_j, such that v_{ij} contains $v_{(i-1)j}$, and $\lim v_{ij}=\alpha_j$. Then by CIII, for i sufficiently large, we have

$$\text{(1)}\qquad L_k(AV_i)>L_k(\alpha' A)-\epsilon>L_k(A)-3\epsilon.$$

Put the part of B exterior to α in β so that any point of β is distant from V_i by not less than $\delta>0$. Then

$$L_k(B-\beta B)\leqq L_k(\alpha B)<\epsilon,$$

and, by CIV,

$$L_k(V_iA) + L_k(\beta A) = L_k(V_iA + \beta A) \leqq L_k(A),$$

which, with (1), gives

$$L_k(\beta A) < 3\epsilon.$$

THEOREM IV. *If A and B are separated sets, both of extent k, then*

$$L_k(A) + L_k(B) = L_k(A + B).$$

Put A in a set of open convex regions α so that

$$L_k(\alpha B) < \epsilon, L_k(A - \alpha A) < \epsilon. \tag{1}$$

Set $E=A+B$. Then, by Lemma II,

$$L_k(\alpha E) + L_k(E - \alpha E) = L_k(E), \tag{2}$$

which, from CI, gives

$$L_k(\alpha A) + L_k(B - \alpha B) \leqq L_k(E). \tag{3}$$

But from Lemma II we get

$$L_k(\alpha A) + L_k(A - \alpha A) = L_k(A), \tag{4}$$

and

$$L_k(\alpha B) + L_k(B - \alpha B) = L_k(B). \tag{5}$$

It then follows from (1), (3), (4), (5), and the fact that ϵ is arbitrary, that

$$L_k(A) + L_k(B) \leqq L_k(A + B).$$

But by CII

$$L_k(A) + L_k(B) \geqq L_k(A + B).$$

These two inequalities give the theorem.

Let $A_B{}^0$ be the part of A for which

$$\overline{\lim_{r\to 0}} \frac{L_k[BH(a, r)]}{L_k[AH(a, r)]} = 0,$$

and $A_B{}^+$ the part of A for which

$$\overline{\lim_{r\to 0}} \frac{L_k[BH(a, r)]}{L_k[AH(a, r)]} > 0.$$

Define $B_A{}^0$ and $B_A{}^+$ by interchanging the roles of A and B.

The ratios which are used in determining the sets A_B^0 and A_B^+ are defined for almost all A. For otherwise there would exist a part of A with outer measure>0 for each point a of which $D^*(a)=0$. But this contradicts Theorem II. Likewise the ratios used in determining the sets B_A^0 and B_A^+ are defined for almost all B.

THEOREM V. *The set A_B^0 is separated from B, and the set B_A^0 is separated from A.*

For a given $\epsilon>0$ there corresponds to each point of A_B^0 a number $\rho>0$ such that

$$\frac{L_k[BH(a, r)]}{L_k[AH(a, r)]} < \epsilon,\ r < \rho. \tag{1}$$

If E is the part of A_B^0 for which (1) holds, then for ρ sufficiently small it follows from CIII that

$$L_k(E) > L_k(A_B^0) - \epsilon. \tag{2}$$

About each point of E there then exists a sequence of hyperspheres $H(e, r_i)$ with r_i tending to zero for which (1) holds. Vitali's argument may now be used to show the existence of a non-overlapping set of these hyperspheres $H=H_1, H_2, \cdots$ which contain almost all of E. By (2) and Lemma II we get

$$L_k(A_B^0 - HA_B^0) < \epsilon,$$

and from (1)

$$\sum L_k(BH_i) < \epsilon \sum L_k(AH_i),$$
$$L_k(BH) < \epsilon L_k(A),$$

where ϵ is arbitrary. Thus A_B^0 is separated from B. In a similar manner it may be shown that B_A^0 is separated from A.

THEOREM VI. *There is no part of A_B^+ with outer measure greater than zero which is separated from B, and no part of B_A^+ with outer measure greater than zero which is separated from A.*

Suppose there is such a part of A_B^+. CIII may then be used to show the existence of a positive number d and a part E of A_B^+ such that at each point e of E we have

$$\frac{L_k[BH(e, r_i)]}{L_k[AH(e, r_i)]} > d \tag{1}$$

for a properly chosen sequence of values of r_i tending to zero. By supposition, E is separated from B. It is, therefore, possible to put a part of E in a set of

open convex regions α in such a way that

$$(2) \qquad L_k(\alpha B) < \epsilon, \text{ and } L_k(E - \alpha E) < \epsilon.$$

About each point of E on α may be put a sequence of hyperspheres $H(e, r_i)$ satisfying (1) and such that all the hyperspheres are on α. Vitali's argument may now be used to show the existence of a countable non-overlapping set of these hyperspheres containing almost all the part of E on α. For this set H_n of hyperspheres we get from (1)

$$\sum L_k(BH_n) > d \sum L_k(AH_n) > dL_k(E) - \epsilon.$$

But this contradicts (2). We conclude, therefore, that there is no part of A_B^+ with outer measure >0 which is separated from B. In a similar manner it can be shown that there is no part of B_A^+ with outer measure >0 which is separated from A.

It has now been shown that A_B^0 is separated from B and no part of A_B^+ is separated from B, with similar remarks applying to B_A^0 and B_A^+. From these facts it is easy, by methods used above, to obtain

THEOREM VII. *A_B^0 is separated from A_B^+ and B_A^0 is separated from B_A^+.*

We next prove

THEOREM VIII. $L_k(A_B^+) = L_k(B_A^+) = L_k(A_B^+ + B_A^+)$.

Suppose $L_k(A_B^+) = L_k(B_A^+) + c$, $c > 0$. Making use of Lemma III, a number $\rho > 0$ may be so fixed that for any set of open convex regions $V = v_1, v_2, \cdots$ with $d_i < \rho$ we have

$$(1) \qquad L_k(VA_B^+) < \sum l_k^i + \frac{c}{4} \cdot$$

Now let V with $d_i < \rho$ enclose B_A^+ in such a way that

$$\sum l_k^i < L_k(B_A^+) + \frac{c}{4} \cdot$$

This, with (1), gives

$$L_k(VA_B^+) < L_k(B_A^+) + \frac{c}{2},$$

which shows that there is a part E of A_B^+ exterior to V with $L_k(E) > c/2$. But V contains B_A^+. Hence, since E is exterior to V, it may be shown by methods used above that E, a part of A_B^+ with outer measure >0, is separated from B. But this contradicts Theorem VI. We conclude, therefore, that $L_k(A_B^+) \leqq L_k(B_A^+)$. Precisely the same argument shows that $L_k(B_A^+) \leqq L_k(A_B^+)$. We thus have $L_k(A_B^+) = L_k(B_A^+)$, which is the first part of the theorem.

Suppose $L_k(A_B^+ + B_A^+) = L_k(A_B^+) + c$, $c > 0$. Lemma III may be used to enclose A_B^+ in a set of open convex regions $V = v_1, v_2, \cdots$ in such a way that

$$\sum l_k^i < L_k(A_B^+) + \frac{c}{4}, \tag{2}$$

and

$$L_k[(A_B^+ + B_A^+)V] < \sum l_k^i + \frac{c}{4}, \tag{3}$$

which shows that there is a part E of $A_B^+ + B_A^+$ exterior to V with measure $> c/2$. Since V contains A_B^+ it follows that E belongs to B_A^+. Reasoning as above, we arrive at the conclusion that E is separated from A_B^+. But this again contradicts Theorem VI. Thus we conclude that $L_k(A_B^+ + B_A^+) \leqq L_k(A_B^+)$. Hence, since always $L_k(A_B^+ + B_A^+) \geqq L_k(A_B^+) = L_k(B_A^+)$, we have

$$L_k(A_B^+ + B_A^+) = L_k(A_B^+) = L_k(B_A^+).$$

From Theorems IV, V, and VII, we get

THEOREM IX.

$$L_k(A) = L_k(A_B^0) + L_k(A_B^+),$$

$$L_k(B) = L_k(B_A^0) + L_k(B_A^+),$$

$$L_k(A + B) = L_k(A_B^0) + L_k(B_A^0) + L_k(A_B^+ + B_A^+).$$

Theorems VIII and IX may now be combined to give

THEOREM X.

$$L_k(A) + L_k(B) = L_k(A + B) + L_k(A_B^+ + B_A^+).$$

5. Relations between sets in general and measurable sets. Let A be any set of finite extent k in S_n, B the complement of A. Let C be the part of B for which

$$\overline{\lim_{r \to 0}} \frac{L_k[AH(c, r)]}{h_k^r} > 0. \tag{1}$$

THEOREM XI. *The set C is of extent not greater than k.*

Suppose there is some integer $j \geqq 1$ for which

$$L_{k+j}(C) > 0.$$

On account of (1) there exist two positive numbers δ and d, and a part C_1 of C with $L_{k+j}(C_1) > 0$ for which

(2) $$\frac{L_k[AH(c_1, r)]}{h_k^r} > d$$

for a proper choice of $r<\delta$. Since $L_{k+j}(C_1)>0$, it follows from Lemma I that there exists a part C_2 of C_1 with $L_k(C_2)>G$, G an arbitrary positive number. Choose a sequence $\delta_1>\delta_2> \cdots$ tending to zero, and let C_2^i be the part of C_2 for which (2) holds for some $r>\delta_i$. Then C_2^i tends to C_2. Thus there exists $\delta'>0$ and a part C_3 of C_2 with $L_k(C_3)>G$ for which we have

(3) $$\frac{L_k[AH(c_3, r)]}{h_k^r} > d$$

for some $r>\delta'$. Now put C_3 in a set of open convex regions $u_1, u_2, \cdots$ with $d_i<\delta'$, and such that

$$\sum l_k^i > L_k(C_3) - \epsilon > G.$$

In each u_i choose a point c_3 of C_3 and about this point put a hypersphere $H(c_3, r)$ with $r_i>\delta'$ and satisfying (3). Then $h_k^i>l_k^i$, and consequently from (3) we get

$$\frac{L_k[AH(c_3, r_i)]}{l_k^{r_i}} \geqq \frac{L_k[AH(c_3, r_i)]}{h_k^{r_i}} > d.$$

This gives

$$\sum L_k[AH(c_3, r_i)] > d \sum l_k^{r_i} > dG.$$

But since $L_k(A)$ is finite, and since G can be chosen arbitrarily large, this gives a contradiction. Hence our assertion is proved.

THEOREM XII. *If A is any set of finite extent k, then a necessary and sufficient condition that A be measurable is that $L_k(C)=0$.*

Let W be any set of extent k. We show that if $L_k(C)=0$, then

(1) $$L_k(W) = L_k(AW) + L_k(W - AW).$$

Set $W-AW=E$. Since $E-EC$ belongs neither to A nor to C, we have for any point e of $E-EC$

$$\overline{\lim_{r\to 0}} \frac{L_k[AH(e, r)]}{h_k^r} = 0.$$

Hence, since $L_k(C)=0$, for almost all E we have

(2) $$\overline{\lim_{r\to 0}} \frac{L_k[AH(e, r)]}{L_k[EH(e, r)]} \Big/ \frac{h_k^r}{L_k[EH(e, r)]} = 0.$$

But for almost all E

$$\overline{\lim_{r\to 0}}\ \frac{h_k^r}{L_k[EH(e, r)]} \geqq 1,$$

and this, with (2), gives, for almost all E,

$$\overline{\lim_{r\to 0}}\ \frac{L_k[AH(e, r)]}{L_k[EH(e, r)]} = 0.$$

Hence almost all $E-EC$ belongs to W_A^0. And since $L_k(EC)=L_k(C)=0$, it follows that almost all E belongs to W_A^0. Hence $E=W-AW$ is separated from A and consequently from WA. The truth of (1) now follows from Theorem IV. Then, according to our definition of measurability, A is measurable. Thus the condition is sufficient.

At every point of C

$$\overline{\lim_{r\to 0}}\ \frac{L_k[AH(c, r)]}{h_k^r} > 0.$$

Hence for almost all C

$$\overline{\lim_{r\to 0}}\ \frac{L_k[AH(c, r)]}{L_k[CH(c, r)]} \Big/ \frac{h_k^r}{L_k[CH(c, r)]} > 0.$$

But for almost all C

$$\underline{\lim_{r\to 0}}\ \frac{h_k^r}{L_k[CH(c, r)]} \geqq 1.$$

Consequently, for almost all C,

$$\overline{\lim_{r\to 0}}\ \frac{L_k[AH(c, r)]}{L_k[CH(c, r)]} > 0.$$

We conclude, therefore, that almost all C belongs to C_A^+. Hence if $L_k(C)>0$ it follows that $L_k(C_A^+)>0$. Now let $W=A_C^+ + C_A^+$. Then by Theorem VIII,

$$L_k(W) = L_k(A_C^+) = L_k(C_A^+).$$

But

$$L_k(AW) + L_k(W - AW) = L_k(A_C^+) + L_k(C_A^+) = 2L_k(W).$$

Hence A is not measurable. This shows that the condition is necessary.

THEOREM XIII. *Let A be any set of finite extent k. Then the set*

$$A' = A + C$$

contains A, is measurable with

$$L_k(A') = L_k(A),$$

and is well-defined in terms of A.

A point b of the set B complementary to the set A does or does not belong to C according as the upper limit of

$$\frac{L_k[AH(b, r)]}{h_k^r}$$

is greater than zero, or is equal to zero. Hence C, and consequently A', is effectively defined in terms of A.

To show that A' is measurable, let c' be a point of C'. Then c' belongs neither to A nor to C, and

$$\overline{\lim_{r\to 0}} \frac{L_k[A'H(c', r)]}{h_k^r} > 0.$$

But this, with Theorems IV, VII, and VIII, gives

$$\lim_{r\to 0} \left\{\frac{L_k[A_{c'}^0 H(c', r)]}{h_k^r} + \frac{L_k[(A_{c'}^+ + C_A^+)H(c', r)]}{h_k^r}\right\} > 0,$$

$$\overline{\lim_{r\to 0}} \frac{L_k[(A_{c'}^0 + A_{c'}^+)H(c', r)]}{h_k^r} > 0,$$

which makes c' a point of C. Hence C' is empty and $L_k(C')=0$. Then by Theorem XII A is measurable.

THEOREM XIV. *If the set A of extent k is regular then A' is regular, and if this set is irregular then A' is irregular.*

Let A be regular, and suppose there is a part of A', other than null parts, at which A' is irregular. Obviously A' is regular at each point of A. In the proof of Theorem XII it was shown that almost all C belongs to C_A^+. Hence there is a part E of C_A^+ for which

$$\frac{L_k[A'H(e, r)]}{h_k^r} < 1 - \eta \tag{1}$$

for an infinite set of arbitrarily small r, $\eta>0$, and $L_k(E)>0$. Then, since E belongs to C_A^+, almost all points of E are points of E_A^+. This and Theorem VIII then give

$$L_k(A_E^+) = L_k(E_A^+) = L_k(E) > 0.$$

The set A' is regular at points of A and consequently at points of A_E^+. Let

F be the part of A_E^+ which is such that

$$\frac{L_k[A'H(f, r)]}{h_k^r} > 1 - \frac{\eta}{2} \tag{2}$$

for $r<\delta$, $L_k(F)>0$. The set F belongs to A_E^+, and consequently almost all F belongs to F_E^+. Hence, as above,

$$L_k(E_F^+) = L_k(F_E^+) = L_k(F) > 0.$$

For each point of E_F^+,(1) holds. Hence about a fixed point x of this set there exists $H(x, r)$ with $r<\delta$ such that

$$\frac{L_k[A'H(x, r)]}{h_k^r} < 1 - \eta. \tag{3}$$

Every point of E_F^+ is a limit point of points of F_E^+. Let $x_1, x_2, \cdots$ be a sequence of points of F_E^+ tending to x. On account of (2) for an arbitrary ϵ there exists $H(x_i, r-\epsilon)$ about each x_i for which

$$\frac{L_k[A'H(x_i, r - \epsilon)]}{h_k^{r-\epsilon}} > 1 - \frac{\eta}{2}\cdot \tag{4}$$

For ϵ fixed and i sufficiently large $H(x_i, r-\epsilon)$ is interior to $H(x, r)$. From this and (4) we have

$$L_k[A'H(x, r)] \geqq L_k[A'H(x_i, r - \epsilon)] > h_k^{r-\epsilon} - \tfrac{1}{2}\eta h_k^{r-\epsilon}. \tag{5}$$

But for ϵ sufficiently small $h_k^{r-\epsilon}$ is arbitrarily near to h_k^r, which makes (5) and (3) contradictory. We conclude, therefore, that if A is regular A' is regular.

The proof for the case when A is irregular is along the same lines, and we merely sketch it. Except for a null set

$$A' = A_C^0 + A_C^+ + C_A^+.$$

Since A_C^0 is separated from C_A^+ it readily follows that A' is irregular at each point of A_C^0. Also, since for any $H(a_c^+, r)$,

$$L_k[(A_C^+ + C_A^+)H(a_c^+, r)] = L_k[A_C^+H(a_c^+, r)],$$

it follows that A' is irregular at each point of A_C^+. Let E be the part of C at which A' is regular. Let F be the part of A_E^+ for which

$$\frac{L_k[A'H(f, r)]}{h_k^r} < 1 - \eta \tag{1}$$

for an infinite set of arbitrarily small r. Let G be the part of E_F^+ at which, for all $r<\delta$,

$$(2) \qquad \frac{L_k[A'H(g, r)]}{h_k^r} > 1 - \frac{\eta}{2}.$$

Then for F_G^+, (1) holds, and for G_F^+, (2) holds. It is now possible to take a point x of F_G^+, and a sequence of points $x_1, x_2, \cdots$ of G_F^+ tending to x, and arrive at a contradiction, as in the case when A was regular.

6. **Some applications.** It was shown in the introduction that corresponding to a set A of linear extent there was a measurable set $\bar{A}$ which contained A, and for which

$$L_k(\bar{A}) = L_k(A).$$

It can likewise be shown that there is a measurable set $\bar{A}$ similarly related to any set A of extent k. We are now in a position to discuss the density properties of this set $\bar{A}$.

If A is regular (irregular) then $\bar{A}$ is regular (irregular).

To prove this, set $\bar{A}=A+B$ where $L_k(B)>0$. If $L_k(B)=0$ the case is trivial. The set B is not separated from A. For then we would have

$$L_k(\bar{A}) = L_k(A) + L_k(B)$$

which cannot hold, since $L_k(\bar{A})=L_k(A)$. Hence almost all B belongs to B_A^+. Now suppose that A is regular but that there is a part of B_A^+ with outer measure greater than zero at which A is irregular. Let E be the part of B_A^+ at which

$$\frac{L_k[\bar{A}H(e, r)]}{h_k^r} < 1 - \eta$$

for a sequence of values of r tending to zero, and $L_k(E)>0$. Since E belongs to B_A^+, each point of E is a point of E_A^+. Let F be the part of A_E^+ which is such that

$$(1) \qquad \frac{L_k[\bar{A}H(f, r)]}{h_k^r} > 1 - \frac{\eta}{2}$$

for $r<\delta$, and $L_k(F)>0$. Each point of F is a point of F_A^+. About a point x of E_F^+ put a hypersphere $H(x, r)$ with $r<\delta$, and such that

$$(2) \qquad \frac{L_k[\bar{A}H(x, r)]}{h_k^r} < 1 - \eta.$$

Let x_i be a sequence of values of F_E^+ tending to x. About each x_i put a hypersphere $H(x_i, r-\epsilon)$. Then, on account of (1), we have

$$(3) \qquad \frac{L_k[\bar{A}H(x_i, r - \epsilon)]}{h_k^{r-\epsilon}} > 1 - \frac{\eta}{2}.$$

For any ϵ, i can be taken large enough to insure that $H(x_i, r-\epsilon)$ is interior to $H(x, r)$. But for η fixed, ϵ can be taken arbitrarily small, which, with (2), (3), and the fact that for ϵ small $h_k^{r-\epsilon}$ is near to h_k^r, leads to a contradiction. We conclude, therefore, that if A is regular then $\overline{A}$ is regular at almost all points. Similar reasoning shows that if A is irregular then $\overline{A}$ is irregular at almost all points.

If A is any plane set of linear extent then the sets $\overline{A}$ and A' are linearly measurable, and are regular or irregular according as A is regular or irregular. Besicovitch has shown that linearly measurable regular plane sets have a tangent at almost all points. Hence if A is regular there exists a tangent at almost all points of $\overline{A}$, and of A'. And since each of these sets contains A it follows that A has a tangent at almost all points. Likewise the other theorems which Besicovitch has proved for linearly measurable plane sets are seen to hold for general sets of linear extent.

In proving that a regular linearly measurable plane set A has a tangent at almost every point, Besicovitch† makes use of the set A_1 which is the part of A for which

$$\left| \frac{L_1[AH(a,r)]}{h_1^r} - 1 \right| \leqq \eta$$

for $r<\delta$. He assumes that A_1 is linearly measurable. The measurability of this set can hardly be considered as obvious, and there seems to be no trivial proof for his assertion. We shall establish some general results from which the measurability of A_1 follows.

We show first that

Separated divisions of measurable sets are measurable.

Let A be any measurable set of extent k, A_1 and A_2 separated divisions of A. Since A_1 and A_2 are separated, C_1 contains at most a null part of A_2, and C_2 contains at most a null part of A_1. Hence, except for at most a null set, C_1 and C_2 belong to C. But since A is measurable, Theorem XII gives $L_k(C)=0$. But this makes $L_k(C_1)=L_k(C_2)=0$, which again by Theorem XII makes A_1 and A_2 measurable.

Let A be any set of extent k. Let B be the part of A for which

$$\frac{L_k[AH(a, r)]}{h_k^r} > 1 - \eta \tag{1}$$

for $r<\delta$. Then the sets B and $E=A-B$ are separated.

† Loc. cit., p. 438.

Suppose B_E^+ exists with $L_k(B_E^+)>0$. Then, by Theorem VIII, $L_k(E_B^+) = L_k(B_E^+)>0$. For each point x of E_B^+ there exists some $r<\delta$ for which

$$\frac{L_k[AH(x, r)]}{h_k^r} \leqq 1 - \eta. \tag{2}$$

Take a sequence of points $x_1, x_2, \cdots$ of B_E^+ tending to x, and about each x_i put a hypersphere $H(x_i, r-\epsilon)$. Then from (1) we have

$$\frac{L_k[AH(x_i, r - \epsilon)]}{h_k^{r-\epsilon}} > 1 - \eta. \tag{3}$$

But, for every ϵ, i can be taken so large that $H(x_i, r-\epsilon)$ is interior to $H(x, r)$. Then, since ϵ can be taken arbitrarily small independent of η, (2) and (3) are contradictory, which allows us to conclude that B and $E=A-B$ are separated.

It can likewise be shown that the part of A for which

$$1 - \eta \leqq \frac{L_k[AH(a, r)]}{h_k^r} \leqq 1 + \eta$$

is separated from the remainder of A. From this it follows that the set A_1 of Besicovitch is separated from $A-A_1$. Then, since A is measurable, A_1 and $A-A_1$ are measurable. We note further that if A is not measurable the sets A_1 and $A-A_1$ are, nevertheless, separated. This fact permits the arguments of Besicovitch in regard to tangency to be carried through for any plane set of linear extent.

ACADIA UNIVERSITY,
WOLFVILLE, NOVA SCOTIA

SUBHARMONIC FUNCTIONS AND MINIMAL SURFACES*

BY

E. F. BECKENBACH† AND TIBOR RADÓ

Introduction

0.1. Let $f(w)$, given by

$$f(w) = x(u, v) + iy(u, v),\ w = u + iv,\ w \text{ in } D,$$

where D is some domain of definition, be an analytic function of the complex variable w. Then $x(u, v)$, $y(u, v)$ satisfy the Cauchy-Riemann differential equations

$$x_u = y_v, \quad x_v = -y_u, \tag{1}$$

the subscripts denoting differentiation. These equations (1) are not symmetric in x, y, but they imply the symmetric set

$$x_u^2 + y_u^2 = x_v^2 + y_v^2,\ x_u x_v + y_u y_v = 0. \tag{2}$$

Conversely, (2) implies either (1) or

$$y_u = x_v,\ y_v = -x_u. \tag{3}$$

From either (1), (2) or (3) it follows that $x(u, v)$, $y(u, v)$ are harmonic functions:

$$x_{uu} + x_{vv} = 0,\ y_{uu} + y_{vv} = 0.$$

If (1) holds, $y(u, v)$ is said to be the conjugate harmonic function of $x(u, v)$, or if (3) holds, $x(u, v)$ is said to be the conjugate harmonic function of $y(u, v)$; generally, if (2) holds then $x(u, v)$, $y(u, v)$ will be called a *couple of conjugate harmonic functions.*

0.2. Generalizing this situation to the case of three functions $x(u, v)$, $y(u, v)$, $z(u, v)$, (u, v) in D, we shall call $x(u, v)$, $y(u, v)$, $z(u, v)$ a *triple of conjugate harmonic functions* provided the following conditions are satisfied:

$$E = G, \quad F = 0, \tag{i}$$

where

$$E = x_u^2 + y_u^2 + z_u^2,\ F = x_u x_v + y_u y_v + z_u z_v,\ G = x_v^2 + y_v^2 + z_v^2;$$

$$x(u, v),\ y(u, v),\ z(u, v) \tag{ii}$$

are harmonic.

* Presented to the Society, December 29, 1932; received by the editors January 23, 1933.

† National Research Fellow.

It might be noted that if one of the coordinate functions vanishes identically, say $z \equiv 0$, then (ii) is implied by (i); but in general this implication does not hold.

0.3. While this generalization no doubt would be of interest from a purely analytic viewpoint, the following *theorem of Weierstrass* shows that it actually is very important geometrically: A necessary and sufficient condition that a surface given in terms of isothermic parameters (that is, parameters u, v such that $E=G$, $F=0$) be minimal is that the coordinate functions be harmonic.

Thus the theory of minimal surfaces appears as the theory of triples of conjugate harmonic functions, while the theory of couples of conjugate harmonic functions is the theory of analytic functions of a complex variable. As a matter of fact, theorems and methods in theory of functions always have served as tools and models in the theory of minimal surfaces.

0.4. The purpose of the present paper is the development of this analogy in the direction of *the principle of the maximum*. If $f(w)$ is an analytic function in a region R, then $|f(w)|$ takes on its maximum on the boundary of R. Similarly, if $x(u, v)$, $y(u, v)$, $z(u, v)$ form a triple of conjugate harmonic functions in R, then $(x^2+y^2+z^2)^{1/2}$ takes on its maximum on the boundary of R; this is easily shown to be true even if the three harmonic functions are not conjugate. However, the effectiveness of the principle of the maximum in the case of analytic functions depends essentially upon the fact that certain operations (multiplication for instance), if performed on analytic functions, yield analytic functions again. This situation does not seem to admit of any direct generalization to minimal surfaces. It is our purpose to show that despite this lack of direct analogy many important applications of the principle of the maximum can be generalized to minimal surfaces. Our tool is the following simple lemma (see §2):

Three functions $x(u, v)$, $y(u, v)$, $z(u, v)$, continuous in a domain, form there a triple of conjugate harmonic functions if and only if $\log[(x+a)^2+(y+b)^2+(z+c)^2]^{1/2}$ *is subharmonic for every choice of the real constants a, b, c.*

This lemma permits us to apply the theory of subharmonic functions,* so important in theory of functions, to the theory of minimal surfaces. For the convenience of the reader, we give in §1 the necessary definitions and facts concerning subharmonic functions.

* See F. Riesz, *Sur les fonctions subharmoniques et leur rapport à la théorie du potentiel* (in two parts), Acta Mathematica, vol. 48 (1926), pp. 329–343, and vol. 54 (1930), pp. 321–360; P. Montel, *Sur les fonctions convexes et les fonctions sousharmoniques*, Journal de Mathématiques, (9), vol. 7 (1928), pp. 29–60; S. Saks, *Sur une inégalité de la théorie des fonctions*, Acta Szeged, vol. 4 (1928), pp. 51–55, and *On subharmonic functions*, Acta Szeged, vol. 5 (1932), pp. 187–193.

1. Subharmonic functions and functions of class PL

1.1. In this section we present the definition of subharmonic functions and give those results concerning these functions which we shall need in the sequel.

Let $g(u, v)$ be a continuous function of two variables, defined in a domain D (connected open set). Suppose that for each point (u_0, v_0) of D we have

$$g(u_0, v_0) \leqq \frac{1}{2\pi}\int_0^{2\pi} g(u_0 + \rho\cos\phi,\ v_0 + \rho\sin\phi)d\phi \tag{4}$$

for each sufficiently small value of the radius ρ. Then the function $g(u, v)$ is said to be *subharmonic* in D.*

The definition can be extended to the case of discontinuous functions, but we shall be concerned in this paper only with continuous subharmonic functions.

1.2. It follows immediately from the definition that a subharmonic function $g(u, v)$ cannot attain its maximum value at any (interior) point of D, unless $g(u, v)$ is identically constant.†

1.3. If a function $g(u, v)$ has continuous partial derivatives of the second order, then a necessary and sufficient condition that $g(u, v)$ be subharmonic is that its Laplacian be $\geqq 0$:

$$\Delta g \equiv g_{uu} + g_{vv} \geqq 0.\ddagger$$

1.4. Let $g(u, v)$ be subharmonic in the ring

$$r_1 < [(u - u_0)^2 + (v - v_0)^2]^{1/2} < r_2,$$

and let $M(r)$ denote the maximum of $g(u, v)$ on

$$(u - u_0)^2 + (v - v_0)^2 = r^2,\ r_1 < r < r_2.$$

Then $M(r)$ is a convex function of $\log r$.§

1.5. Obviously, if $g(u, v)$ and $h(u, v)$ are both subharmonic in D, then $g(u, v) + h(u, v)$ also is subharmonic there.

---

* This definition is due to F. Riesz. See Acta Mathematica, loc. cit., first part, p. 331.

† See F. Riesz, Acta Mathematica, loc. cit., first part, p. 331.

‡ See F. Riesz, Acta Mathematica, loc. cit., first part, p. 335.

§ See P. Montel, Journal de Mathématiques, loc. cit., where this fact and similar elementary facts concerning subharmonic functions are presented in a systematic way.

1.6. A function $p(u, v)$, defined in a domain D, will be said to be *of class* PL in D provided the following conditions are satisfied there.

(i) $p(u, v)$ is continuous.

(ii)
$$p(u, v) \geqq 0.$$

(iii) $\log p(u, v)$ is subharmonic in the part of D where $p(u, v) > 0$.

1.7. If $p(u, v)$ is of class PL, then $p(u, v)$ is subharmonic. Indeed, at points where $p(u, v) = 0$ the condition (4) of Riesz obviously is satisfied; and elsewhere the fact that $\log p(u, v)$ is subharmonic implies that $p(u, v)$ is subharmonic.*

1.8. Obviously (see §1.5), the product of a finite number of functions of class PL, or any positive power of a function of this class, is again a function of class PL.

1.9. The class PL is invariant under conformal mapping. (The same remark applies to the class of subharmonic functions.) That is, if $p(u, v)$ is of class PL in D and if D is mapped conformally on a (U, V) domain $\overline{D}$, then $p(u, v)$ is transformed into a function $q(U, V)$ which is of class PL in $\overline{D}$.

1.10. A necessary and sufficient condition that a non-negative function $p(u, v)$ be of class PL is that $e^{\alpha u + \beta v}\, p(u, v)$ be subharmonic for every choice of the real constants α, β.† It follows from this (see §1.5) that the sum of a finite number of functions of class PL is again a function of class PL.

1.11. The classical example of a function of class PL is the absolute value of an analytic function $f(w)$ of $w = u + iv$. If $f(w)$ is different from zero in a domain, then $\log |f(w)|$ is harmonic there. Thus $|f(w)|$ is just barely of class PL. As a consequence, a great number of theorems concerned with $|f(w)|$ are a fortiori true for functions of class PL. We now shall state some of these generalized theorems which will be used in the sequel. The proofs run exactly in the same way as for $|f(w)|$; for this reason we shall sketch just a few of the proofs, and otherwise shall give references to typical proofs concerning $|f(w)|$.

1.12. Let $p(u, v)$ be bounded and of class PL in $u^2 + v^2 < 1$. Suppose $p(u, v)$ remains continuous on a certain arc σ of $u^2 + v^2 = 1$, and vanishes there. Then $p(u, v) \equiv 0$.

Proof.‡ Choose the integer n so large that $2\pi/n$ is less than the length of

* See P. Montel, Journal de Mathématiques, loc. cit., p. 39.

† This criterion is due to Montel, Journal de Mathématiques, loc. cit., p. 40, who proved it under the assumption that $p(u, v)$ has continuous partial derivatives of the first and second order. For the case of a merely continuous $p(u, v)$, the theorem has been proved by T. Radó, *Remarque sur les fonctions subharmoniques*, Paris Comptes Rendus, vol. 186 (1928), pp. 346–348.

‡ Cf. Pólya und Szegö, *Aufgaben und Lehrsätze aus der Analysis*, Berlin, J. Springer, 1925, vol. I, p. 139, problem 279.

the arc σ. If we rotate the unit circle about its center through an angle of $2\pi/n$, $p(u, v)$ is transformed into a new function $p_1(u, v)$ of class PL (see §1.9). Let $p_2(u, v), \cdots, p_{n-1}(u, v)$ be the functions of class PL resulting from further successive rotations of the unit circle through the angle $2\pi/n$. Then $\psi(u, v) = pp_1 \cdots p_{n-1}$ is again of class PL (see §1.8), and $\psi(u, v) \to 0$ if (u, v) converges to any point of $u^2+v^2=1$. Since $\psi \geqq 0$, it follows from this (see §1.2) that $\psi(u, v) \equiv 0$. In particular, $\psi(0, 0) = p(0, 0)^n = 0$, that is to say, $p(u, v)$ vanishes at the origin. As any point of $u^2+v^2<1$ can be thrown, by conformal mapping of the unit circle upon itself, into the origin, it follows that $p(u, v) \equiv 0$.

1.13. Let $p(u, v)$ be bounded and of class PL in $u^2+v^2<1$. Suppose $p(u, v)$ vanishes in a subdomain k of $u^2+v^2<1$. Then $p(u, v) \equiv 0$.

Proof. Consider any fixed point (u_0, v_0) of k. Then given any point (u_1, v_1) in $u^2+v^2<1$ but not in k, there exists a circle passing through (u_0, v_0), tangent to $u^2+v^2=1$ from within, and containing (u_1, v_1) in its interior. The theorem of §1.12 applies to this circle.

1.14. Let $p(u, v)$ be bounded and of class PL in the angle $0 < \text{arc tg } (v/u) < \alpha$. Let $p(u, v)$ remain continuous on the ray $u>0$, $v=0$, and let $p(u, 0) \to 0$ as $u \to +0$. Then in every angle $0 < \text{arc tg } (v/u) < \alpha - \sigma$, where $\sigma > 0$, we have $p(u, v) \to 0$ as $(u, v) \to (0, 0)$ in any manner.*

Of course this theorem is true if the domain of definition is only the sector $0 < \text{arc tg } (v/u) < \alpha$, $0 < u^2+v^2 < r_0^2$; the proof is the same in either case.

1.15. Let $p(u, v)$ be bounded and of class PL in $u^2+v^2<1$. Let (u', v'), (u'', v'') be two distinct points on $u^2+v^2=1$. Let (u_n', v_n'), (u_n'', v_n'') be two sequences in $u^2+v^2<1$, converging to (u', v'), (u'', v'') respectively, and let C_n be a continuous arc, joining (u_n', v_n') and (u_n'', v_n''), and comprised in the ring $1-\epsilon_n < (u^2+v^2)^{1/2} < 1$, where $\epsilon_n > 0$, and $\epsilon_n \to 0$. Denote by η_n the maximum of $p(u, v)$ on C_n and suppose that $\eta_n \to 0$. Then $p(u, v) \equiv 0$.†

1.16. Let $p(u, v)$ be $\leqq 1$ and of class PL in $r^2 = u^2+v^2 < 1$. Let $p(0, 0) = 0$ and suppose that for a certain $\alpha > 0$, $p(u, v)/r^\alpha$ remains bounded in $0<r<1$. Then $p(u, v) \leqq r^\alpha$. If the equality holds for any (u, v), $0<u^2+v^2<1$, then it holds identically.‡

Proof. Let $M(r)$ denote the maximum of $p(u, v)$ on $u^2+v^2=r^2$. Then

* This generalizes a theorem of Lindelöf. Cf. Pólya und Szegö, loc. cit., p. 138, problem 277. The proof, given there for the special case when $p(u, v)$ is the absolute value of an analytic function, applies without the change of a word to the general case considered above.

† Cf. L. Bieberbach, *Lehrbuch der Funktionentheorie*, Berlin, B. G. Teubner, 1927, vol. II, pp. 19–21.

‡ This generalizes the Lemma of Schwarz. See C. Carathéodory, *Conformal Representation*, London, Cambridge University Press, 1932, p. 39. The example $p(u, v) = (u^2+v^2)^{1/4}$ shows that the value $\alpha = 1$ which holds for the Lemma of Schwarz does not hold in the general case.

$M(r)/r^\alpha$ is the maximum of $p(u, v)/r^\alpha$ on $u^2+v^2=r^2$. Since $\log r^\alpha$ is harmonic, $p(u, v)/r^\alpha$ is of class PL in $0<u^2+v^2<1$. Therefore (see §1.4), $M(r)/r^\alpha$ is a convex function of $\log r$, $-\infty<\log r<0$. If such a function is bounded from above, then it is a non-decreasing function. Consequently, from

$$\lim_{r\to 1} M(r)/r^\alpha \leqq 1$$

it follows that $p(u, v)/r^\alpha \leqq 1$, $0<r<1$. If the equality holds for any (u, v), $0<u^2+v^2<1$, then (see §1.2) it holds identically.

2. A characterization of minimal surfaces

2.1. If $x(u, v)$, $y(u, v)$, $z(u, v)$ form a triple of conjugate harmonic functions (see §0.2) in a domain D, then we shall say that the equations

$$x = x(u, v),\ y = y(u, v),\ z = z(u, v),\ (u, v) \text{ in } D, \tag{5}$$

give a *minimal surface in typical representation*. In this statement, the term *minimal surface* is used in a more general sense than is customary in differential geometry, where the condition $EG-F^2>0$ is always required. In §4.1, we shall use the term minimal surface in an (apparently) even more general sense.

If the equations (5) give a minimal surface $\mathfrak{M}$ in typical representation, then the function $(x^2+y^2+z^2)^{1/2}$ will be called the *norm* of $\mathfrak{M}$ and will be denoted by $|\mathfrak{M}|$ or $|\mathfrak{M}(u, v)|$ or $|\mathfrak{M}(w)|$, where $w=u+iv$.

2.2. *Let*

$$\mathfrak{M}:\quad x = x(u, v),\ y = y(u, v),\ z = z(u, v),\ (u, v) \textit{ in } D, \tag{6}$$

be a minimal surface given in typical representation. Then

$$|\mathfrak{M}| \equiv (x^2 + y^2 + z^2)^{1/2} \tag{7}$$

is of class PL.

It is sufficient to consider points where $|\mathfrak{M}|\neq 0$ (see §1.6). At such points the Laplacian of $\log|\mathfrak{M}|$ is given by $\Delta \log |\mathfrak{M}| = T/|\mathfrak{M}|^4$, with

$$T = (\mathfrak{x}_u^2 + \mathfrak{x}_v^2)\mathfrak{x}^2 - 2[(\mathfrak{x}\mathfrak{x}_u)^2 + (\mathfrak{x}\mathfrak{x}_v)^2],$$

where $\mathfrak{x}$, $\mathfrak{x}_u$, $\mathfrak{x}_v$ denote vectors, namely

$$\mathfrak{x} = (x, y, z),\ \mathfrak{x}_u = (x_u, y_u, z_u),\ \mathfrak{x}_v = (x_v, y_v, z_v),$$

and where the vector products indicated are scalar. The parameters being isothermic, we have

$$\mathfrak{x}_u^2 = \mathfrak{x}_v^2 = \lambda,\ \mathfrak{x}_u\mathfrak{x}_v = 0.$$

Since the partial derivatives of the second order of $\log|\mathfrak{M}|$ are continuous

where $|\mathfrak{M}|\neq 0$, we have only to show (see §1.3) that

$$T \geqq 0. \tag{8}$$

Fix (u_0, v_0); then two cases are possible; either $\lambda=0$ or $\lambda>0$. If $\lambda=0$, then $\mathfrak{x}_u=\mathfrak{x}_v=0$ and (8) is trivial. If $\lambda>0$, then the vectors $\mathfrak{x}_u$, $\mathfrak{x}_v$ are both $\neq 0$ and are perpendicular to each other; let ξ denote the unit vector perpendicular to each of them. Then we can write

$$\mathfrak{x} = a\mathfrak{x}_u + b\mathfrak{x}_v + c\xi,$$

where a, b, c are scalars. Therefore

$$\mathfrak{x}^2 = a^2\lambda + b^2\lambda + c^2,$$
$$\mathfrak{x}\mathfrak{x}_u = a\lambda,\ \mathfrak{x}\mathfrak{x}_v = b\lambda,$$
$$T = 2\lambda(a^2\lambda + b^2\lambda + c^2) - 2(a^2\lambda^2 + b^2\lambda^2) = 2\lambda c^2 \geqq 0.$$

2.3. The fact that (7) is of class PL certainly does not characterize minimal surfaces.* However, (6) is still a minimal surface given in typical representation if we shift the xyz-axes. Therefore, for the functions $x(u, v)$, $y(u, v)$, $z(u, v)$ in (6),

$$[(x + a)^2 + (y + b)^2 + (z + c)^2]^{1/2}$$

is of class PL for arbitrary choice of the real constants a, b, c. And, as we now shall show, the converse also is true, so that we have the following

LEMMA. *A necessary and sufficient condition that the continuous functions $x(u, v)$, $y(u, v)$, $z(u, v)$ represent a minimal surface given in typical representation is that $[(x+a)^2+(y+b)^2+(z+c)^2]^{1/2}$ be of class PL for arbitrary choice of the real constants a, b, c.*

2.4. The necessity has been proved above. To prove the sufficiency, observe first that if $(x^2+y^2+z^2)^{1/2}$ is of class PL, then $x^2+y^2+z^2$ also is of class PL (see §1.8). Let then (u_0, v_0) be any fixed point of D, and put $x(u_0, v_0)=x_0$, $y(u_0, v_0)=y_0$, $z(u_0, v_0)=z_0$. Then if C denotes a sufficiently small circle with center at (u_0, v_0) we have

$$(x_0 + a)^2 + (y_0 + b)^2 + (z_0 + c)^2$$
$$\leqq \frac{1}{2\pi}\int_C [(x + a)^2 + (y + b)^2 + (z + c)^2]d\phi,$$

whence

* See §2.6.

$$
\begin{aligned}
0 \geqq x_0^2 + y_0^2 + z_0^2 - \frac{1}{2\pi}\int_C (x^2+y^2+z^2)d\phi - 2\Big[a\Big(\frac{1}{2\pi}\int_C x d\phi - x_0\Big) \\
+ b\Big(\frac{1}{2\pi}\int_C y d\phi - y_0\Big) + c\Big(\frac{1}{2\pi}\int_C z d\phi - z_0\Big)\Big].
\end{aligned} \tag{9}
$$

The point (u_0, v_0) and the circle C being fixed, the right-hand member of this inequality is a linear function of the arbitrary real constants a, b, c. Thus (9) clearly implies that the coefficients of a, b, c vanish. That is to say, $x(u, v)$ for instance has the property that, for every point (u_0, v_0) in D,

$$
x(u_0, v_0) = \frac{1}{2\pi}\int_0^{2\pi} x(u_0 + \rho\cos\phi, v_0 + \rho\sin\phi)d\phi,
$$

for sufficiently small values of ρ. As is well known, this property characterizes harmonic functions.* Thus it follows that $x(u, v)$, $y(u, v)$, $z(u, v)$ are harmonic functions.

2.5. We proceed to show that $E=G$, $F=0$. Let $\mathfrak{x}=(x, y, z)$, and let $\mathfrak{y}=\mathfrak{x}+\mathfrak{a}$, where $\mathfrak{a}$ is an arbitrary constant vector. By assumption, then, $(\mathfrak{y}^2)^{1/2}$ is of class PL so that (see §§1.3 and 2.2)

$$
(\mathfrak{y}_u^2 + \mathfrak{y}_v^2)\mathfrak{y}^2 - 2[(\mathfrak{y}\mathfrak{y}_u)^2 + (\mathfrak{y}\mathfrak{y}_v)^2] \geqq 0 \tag{10}
$$

at points where $\mathfrak{y}\neq 0$. At points where $\mathfrak{y}=0$, (10) clearly also holds (with the sign of equality).

Consider a definite point (u_0, v_0) in D. Then $\mathfrak{y}_u=\mathfrak{x}_u$, $\mathfrak{y}_v=\mathfrak{x}_v$, regardless of the choice of the constant vector $\mathfrak{a}$. Choose first $\mathfrak{a}=\mathfrak{x}_u(u_0, v_0)-\mathfrak{x}(u_0, v_0)$. Then $\mathfrak{y}(u_0, v_0)=\mathfrak{x}_u(u_0, v_0)$, and (10) gives that

$$
EG - E^2 - 2F^2 \geqq 0
$$

at the point (u_0, v_0). Choose secondly $\mathfrak{a}=\mathfrak{x}_v(u_0, v_0)-\mathfrak{x}(u_0, v_0)$. (10) gives

$$
EG - G^2 - 2F^2 \geqq 0
$$

at (u_0, v_0). Addition gives

$$
-(E-G)^2 - 4F^2 \geqq 0
$$

and consequently $E=G$, $F=0$ at (u_0, v_0). Since (u_0, v_0) was any point in D, the lemma of §2.3 is proved.

2.6. The following remark might help explain the situation.

---

* See for instance O. D. Kellogg, *Foundations of Potential Theory*, Berlin, J. Springer, 1929, p. 227.

*If $\mathfrak{x}=\mathfrak{x}(u, v)$, (u, v) in D, is a minimal surface in typical representation, then $E=\mathfrak{x}_u^2$ is of class PL in D.**

It is clearly sufficient to consider the case when D is the interior of a circle. Then the components $x(u, v)$, $y(u, v)$, $z(u, v)$, which are harmonic, can be written in the form

$$x = \Re f_1(w),\ y = \Re f_2(w),\ z = \Re f_3(w),$$

where $f_1(w)$, $f_2(w)$, $f_3(w)$ are single-valued analytic functions of $w=u+iv$ in D. We have then

$$x_u - ix_v = f_1',\ y_u - iy_v = f_2',\ z_u - iz_v = f_3',$$

and hence, on account of $E=G$,

$$E = \tfrac{1}{2}(|f_1'|^2 + |f_2'|^2 + |f_3'|^2).$$

Thus E is the sum of three functions of class PL, and consequently (see §1.10) E is also of class PL.

As an example, let us consider the surface of Enneper† (in typical representation)

$$\begin{aligned} x &= 3u + 3uv^2 - u^3, \\ y &= -3v - 3u^2v + v^3, \\ z &= 3u^2 - 3v^2. \end{aligned}$$

Then x_u, y_u, z_u are three harmonic functions, such that the sum of their squares is of class PL. Computation shows that x_u, y_u, z_u are not conjugate. Thus, *in the lemma of* §2.3, *the parameters a, b, c are actually necessary, even if the given three functions are known to be harmonic.*

3. Applications

3.1. Let

$$\mathfrak{M}:\quad x = x(u, v),\ y = y(u, v),\ z = z(u, v),\ u + iv = w,\ |w| < 1,$$

be a minimal surface given in typical representation, such that $(0, 0)$ is carried into $(0, 0, 0)$. If $\mathfrak{M}$ is comprised in the unit sphere, $x^2+y^2+z^2\leqq 1$, then

$$|\mathfrak{M}(w)| \leqq |w|,\ 0 < |w| < 1, \tag{11}$$

and

$$E_0^{1/2} \leqq 1, \tag{12}$$

* In a subsequent paper, *Subharmonic functions and surfaces of negative curvature*, in the present number of these Transactions, we point out that if a surface is given in typical representation, then $E=\mathfrak{x}_u^2$ is of class PL if and only if the Gauss curvature of the surface is $\leqq 0$.

† See G. Darboux, *Théorie Générale des Surfaces*, Paris, 1887, vol. I, pp. 372–376.

where $E_0^{1/2}$ denotes the length deformation ratio at the origin. The equalities hold if and only if $\mathfrak{M}$ is a simply-covered circular disc with unit radius.*

Proof. Since $E=G$, $F=0$, we have

$$\lim_{w\to 0} |\mathfrak{M}(w)|/|w| = E_0^{1/2} \tag{13}$$

and therefore $|\mathfrak{M}(w)|/|w|$ remains bounded in $0<|w|<1$. Consequently in $0<|w|<1$ we can apply §1.16, with $\alpha=1$, to the function $p(u, v)=|\mathfrak{M}(w)|$. This gives (11), and then (13) yields (12).

If we define $|\mathfrak{M}(w)|/|w|=E_0^{1/2}$ for $w=0$, then both (11) and (12) are contained in $|\mathfrak{M}(w)|/|w|\leqq 1$, $|w|<1$. If then $|\mathfrak{M}(w)|/|w|=1$ for any w in $|w|<1$, then (see §1.2) the equality is an identity, $|\mathfrak{M}|^2\equiv u^2+v^2$. Differentiation gives

$$\mathfrak{x}\mathfrak{x}_u = u, \mathfrak{x}\mathfrak{x}_v = v,$$

$$\mathfrak{x}\mathfrak{x}_{uu} + \mathfrak{x}_u^2 = 1, \mathfrak{x}\mathfrak{x}_{vv} + \mathfrak{x}_v^2 = 1,$$

whence addition gives $E=G=1$ throughout. Therefore the area of the minimal surface is

$$A = \iint_{u^2+v^2<1} (EG-F^2)^{1/2} du dv = \pi.$$

It follows from this situation that $\mathfrak{M}$ is a simply-covered circular disc.†

3.2. Let

$$\mathfrak{M}: \quad x = x(u, v), y = y(u, v), z = z(u, v), u^2 + v^2 < 1,$$

be a minimal surface given in typical representation, and let $|\mathfrak{M}|$ be bounded. Suppose $x(u, v)$, $y(u, v)$, $z(u, v)$ remain continuous on a certain arc σ of $u^2+v^2=1$, and $x(u, v)=\text{const.}=x_0$, $y(u, v)=\text{const.}=y_0$, $z(u, v)=\text{const.}=z_0$ there. Then $x(u, v)\equiv x_0$, $y(u, v)\equiv y_0$, $z(u, v)\equiv z_0$.‡

Proof. Apply §1.12 to the function

$$p(u, v) = [(x(u, v) - x_0)^2 + (y(u, v) - y_0)^2 + (z(u, v) - z_0)^2]^{1/2}.$$

3.3. Let

$$\mathfrak{M}: \quad x = x(u, v), y = y(u, v), z = z(u, v), 0 < \text{arc tg}(v/u) < \alpha,$$

---

* This generalizes the Lemma of Schwarz. Cf. C. Carathéodory, *Conformal Representation*, p. 39.

† See E. F. Beckenbach, *The area and boundary of minimal surfaces*, Annals of Mathematics, (2), vol. 33 (1932), pp. 658–664.

‡ See T. Radó, *Some remarks on the problem of Plateau*, Proceedings of the National Academy of Sciences, vol. 16 (1930), pp. 242–248; J. Douglas, *Solution of the problem of Plateau*, these Transactions, vol. 33 (1931), pp. 262–321.

be a minimal surface given in typical representation, and let $|\mathfrak{M}|$ be bounded. Let further $x(u, v)$, $y(u, v)$, $z(u, v)$ remain continuous on the ray $u>0$, $v=0$, and let $x(u, 0)\to x_0$, $y(u, 0)\to y_0$, $z(u, 0)\to z_0$ as $u\to +0$. Then in every angle

$$0 < \operatorname{arc\,tg} \frac{v}{u} < \alpha - \sigma, \text{ where } \sigma > 0,$$

we have $x(u, v)\to x_0$, $y(u, v)\to y_0$, $z(u, v)\to z_0$ as $(u, v)\to(0, 0)$ in any manner.*

Proof. Apply §1.14 to the function

$$p(u, v) = [(x(u, v) - x_0)^2 + (y(u, v) - y_0)^2 + (z(u, v) - z_0)^2]^{1/2}.$$

As in §1.14, the theorem is true if the domain of definition is only the sector $0<\operatorname{arc\,tg}(v/u)<\alpha$, $0<u^2+v^2<r_0^2$.

3.4. Besides the assumptions of §3.3, suppose $x(u, v)$, $y(u, v)$, $z(u, v)$ remain continuous on the ray $\operatorname{arc\,tg}(v/u)=\alpha$, $u^2+v^2>0$, and let $x(u, v)\to x_1$, $y(u, v)\to y_1$, $z(u, v)\to z_1$ as $(u, v)\to(0, 0)$ along the ray $\operatorname{arc\,tg}(v/u)=\alpha$. Then $x_0=x_1$, $y_0=y_1$, $z_0=z_1$, and $x(u, v)\to x_0=x_1$, $y(u, v)\to y_0=y_1$, $z(u, v)\to z_0=z_1$ as $(u, v)\to(0, 0)$ in any manner in the angle $0<\operatorname{arc\,tg}(v/u)<\alpha$.

Proof. Apply §3.3 to the angles

$$0 < \operatorname{arc\,tg} \frac{v}{u} < \frac{3\alpha}{4} \text{ and } \frac{\alpha}{4} < \operatorname{arc\,tg} \frac{v}{u} < \alpha$$

and compare results. As before, the theorem is still true if the domain of definition is only the sector

$$0 < \operatorname{arc\,tg} \frac{v}{u} < \alpha, \; 0 < u^2 + v^2 < r_0^2.$$

3.5. The preceding result yields a new proof of the following lemma, used by J. Douglas in his work on the problem of Plateau.†

Let the integrable functions $\xi(\phi)$, $\eta(\phi)$, $\zeta(\phi)$, substituted in the Poisson integral formula, determine the (harmonic) coordinate functions of a minimal surface

$$\mathfrak{M}\colon\; x = x(u, v),\; y = y(u, v),\; z = z(u, v),\; u^2 + v^2 < 1,$$

in typical representation. Let further $\xi(\phi)$, $\eta(\phi)$, $\zeta(\phi)$ approach definite limit values $\xi_-(\pi)$, $\eta_-(\pi)$, $\zeta_-(\pi)$ and $\xi_+(\pi)$, $\eta_+(\pi)$, $\zeta_+(\pi)$ according as $\phi\to\pi$ in clockwise and counterclockwise senses respectively. Then

$$\xi_-(\pi) = \xi_+(\pi),\; \eta_-(\pi) = \eta_+(\pi),\; \zeta_-(\pi) = \zeta_+(\pi). \tag{14}$$

* This generalizes a theorem of Lindelöf. Cf. Pólya und Szegö, loc. cit.

† J. Douglas, loc. cit., pp. 304–306.

Proof. It is a well known property of the Poisson integral that, because of the specified nature of the discontinuity of $\xi(\phi)$ at $\phi=\pi$, the function $x(u, v)$ approaches a definite limit if $(u, v)\to(-1, 0)$ along any straight line in $u^2+v^2<1$, this limit being a linear function of the angle from the u-axis to the straight line and varying from $\xi_-(\pi)$ to $\xi_+(\pi)$ as the angle varies from $-\pi/2$ to $\pi/2$. Similar statements hold for $y(u, v)$, $z(u, v)$. But if we join two such straight lines by a circular arc lying in $u^2+v^2<1$, we obtain a sector for which §3.4 applies; consequently, $(x, y, z)\to$ a definite (x_0, y_0, z_0) which does not vary with the angle. That is, the linear functions mentioned above are constants, whence (14).

3.6. Let

$$\mathfrak{M}: \quad x = x(u, v),\ y = y(u, v),\ z = z(u, v),\ (u, v) \text{ interior to } R,$$

where R is a Jordan region,* be a minimal surface given in typical representation, and let $|\mathfrak{M}|$ be bounded. Let further $x(u, v)$, $y(u, v)$, $z(u, v)$ remain continuous on the boundary of R except possibly at a single point (u_0, v_0), and let $(x, y, z)\to(x_0, y_0, z_0)$ and $(x, y, z)\to(x_1, y_1, z_1)$ as (u, v) converges on the boundary to (u_0, v_0) from one side and the other respectively. Then $(x_0, y_0, z_0)=(x_1, y_1, z_1)$ and $x(u, v)\to x_0=x_1$, $y(u, v)\to y_0=y_1$, $z(u, v)\to z_0=z_1$ as $(u, v)\to(u_0, v_0)$ in any manner in R.

The proof follows immediately from §3.4 by conformal mapping. It can be obtained also by following step by step the proof, for the absolute value of an analytic function of a complex variable, based on the rotation-method.†

4. On conformal maps of minimal surfaces

4.1. The most general definition (actually used in the literature) of a minimal surface is as follows.‡

A set of equations

$$x = \xi(\alpha, \beta),\ y = \eta(\alpha, \beta),\ z = \zeta(\alpha, \beta),\ (\alpha, \beta) \text{ in } R, \tag{15}$$

where R denotes a Jordan region,* defines a *continuous surface S of the topological type of the circular disc*, if $\xi(\alpha, \beta)$, $\eta(\alpha, \beta)$, $\zeta(\alpha, \beta)$ are continuous in R.

The surface (15) is a minimal surface if the following condition is satisfied. Given any point (α_0, β_0) interior to R, there exists a vicinity V_0 of (α_0, β_0) and a topological transformation $\bar\alpha=\bar\alpha(\alpha, \beta)$, $\bar\beta=\bar\beta(\alpha, \beta)$ of V_0, such that $\xi(\alpha, \beta)$, $\eta(\alpha, \beta)$, $\zeta(\alpha, \beta)$ are transformed into functions $\bar\xi(\bar\alpha, \bar\beta)$, $\bar\eta(\bar\alpha, \bar\beta)$, $\bar\zeta(\bar\alpha, \bar\beta)$ which form a triple of conjugate harmonic functions in the image $\overline{V}_0$ of V (see §0.2). Such parameters $\bar\alpha$, $\bar\beta$ are called *local typical parameters*.

* That is, the set of points in and on a Jordan curve.

† Cf. C. Carathéodory, *Conformal Representation*, pp. 21–24.

‡ See T. Radó, *Contributions to the theory of minimal surfaces*, Acta Szeged, vol. 9 (1932), p. 9.

4.2. According to the fundamental theorem in the theory of uniformization,† a minimal surface in the general sense defined above admits also of *typical parameters in the large*, in the following sense. If

$$(16) \qquad S: \quad x = \xi(\alpha, \beta),\ y = \eta(\alpha, \beta),\ z = \zeta(\alpha, \beta),\ (\alpha, \beta) \text{ in } R,$$

is a minimal surface, in the sense of §4.1, then there exists a topological transformation

$$(17) \qquad T: \begin{cases} u = u(\alpha, \beta), & v = v(\alpha, \beta), \quad (\alpha, \beta) \text{ interior to } R, \\ \alpha = \alpha(u, v), & \beta = \beta(u, v), \quad u^2 + v^2 < 1, \end{cases}$$

of the interior of R into $u^2+v^2<1$, such that $\xi(\alpha, \beta)$, $\eta(\alpha, \beta)$, $\zeta(\alpha, \beta)$ are carried into three functions

$$(18) \qquad \begin{aligned} &x(u, v) = \xi(\alpha(u, v), \beta(u, v)),\ y(u, v) = \eta(\alpha(u, v), \beta(u, v)), \\ &z(u, v) = \zeta(\alpha(u, v), \beta(u, v)) \end{aligned}$$

which form a triple of conjugate harmonic functions in $u^2+v^2<1$. Our purpose in this section is to study the situation on the boundary.

4.3. Using the same notations as in the preceding paragraph, §4.2, *suppose that the functions* $\xi(\alpha, \beta)$, $\eta(\alpha, \beta)$, $\zeta(\alpha, \beta)$ *in* (16) *do not all three reduce to constants on any arc of the boundary of* R.

Then the transformation (17) *remains continuous and one-to-one on the boundaries. As a consequence, the functions* $x(u, v)$, $y(u, v)$, $z(u, v)$ *in* (18) *remain continuous on* $u^2+v^2=1$.

4.4. The preceding assertion will be established if we disprove the following two possibilities.

(i) Suppose there exist in the interior of R two sequences (α_n', β_n'), (α_n'', β_n'') converging to the same point (α_0, β_0) on the boundary of R, such that the corresponding sequences (u_n', v_n'), (u_n'', v_n'') converge to two distinct points (u_0', v_0'), (u_0'', v_0'') on $u^2+v^2=1$. Denote then by l_n an arc in the interior of R, connecting (α_n', β_n') and (α_n'', β_n''), such that l_n converges to (α_0, β_0); and denote by C_n the image of l_n in $u^2+v^2<1$. Then the theorem of §1.15 applies to the function

$$p(u, v) = [(x(u, v) - \xi(\alpha_0, \beta_0))^2 + (y(u, v) - \eta(\alpha_0, \beta_0))^2 + (z(u, v) - \zeta(\alpha_0, \beta_0))^2]^{1/2},$$

and it follows that $p(u, v)$ vanishes identically. Hence $x(u, v)$, $y(u, v)$, $z(u, v)$ and consequently $\xi(\alpha, \beta)$, $\eta(\alpha, \beta)$, $\zeta(\alpha, \beta)$ all reduce to constants. This contradicts the assumption stated in §4.3.

(ii) Denote by (α_1, β_1), (α_2, β_2) any two distinct points on the boundary of

† See C. Carathéodory, *Conformal Representation*, chapter VII, and also the bibliographical notes given there on p. 105.

R, and by C a Jordan arc in the interior of R connecting (α_1, β_1) and (α_2, β_2). On account of the preceding result, the image C^* of C is a Jordan arc in $u^2+v^2<1$ with definite end points on $u^2+v^2=1$. We have to disprove the possibility that these end points coincide. Suppose they do coincide. Then C^* is actually a closed Jordan curve, which has a unique point (u_0, v_0) in common with $u^2+v^2=1$. Denote by D^* the interior of C^*. Then $x(u, v)$, $y(u, v)$, $z(u, v)$ satisfy in D^* the assumptions of §3.6. Hence $x(u, v)$, $y(u, v)$, $z(u, v)$ converge to definite limits x_0, y_0, z_0 if (u, v) converges to (u_0, v_0) from within D^*.

D^* is the image of a domain D in R which is bounded by C and by a certain arc σ of the boundary of R. If (α, β) converges, from within D, to any point of σ, then (u, v) converges to (u_0, v_0) from within D^*. Hence

$$\xi(\alpha, \beta) = x(u, v) \rightarrow x_0, \ \eta(\alpha, \beta) = y(u, v) \rightarrow y_0, \ \zeta(\alpha, \beta) = z(u, v) \rightarrow z_0.$$

That is to say, $\xi(\alpha, \beta)$, $\eta(\alpha, \beta)$, $\zeta(\alpha, \beta)$ all three reduce to constants on σ, in contradiction with the assumption made in §4.3.

4.5. We mention the following two special cases of the theorem of §4.3. Suppose that $\xi(\alpha, \beta)\equiv\alpha$, $\eta(\alpha, \beta)\equiv\beta$, $\zeta(\alpha, \beta)\equiv 0$ in the Jordan region R. Then the assumptions of §§4.2 and 4.3 obviously are satisfied and the theorem of §4.3 reduces to the so-called *Osgood-Carathéodory theorem*: If the interior of a Jordan region R is mapped in a one-to-one and conformal way upon $u^2+v^2<1$, the map remains continuous and one-to-one on the boundary of R.†

4.6. Suppose next that the equations (16) carry the boundary of R in a topological way into a Jordan curve Γ. In this case we say that the surface S is bounded by Γ. The theorem of §4.3 implies then the following result. *A minimal surface S (in the general sense of* §4.1), *bounded by a Jordan curve Γ, admits of a representation*

$$x = x(u, v), \ y = y(u, v), \ z = z(u, v), \ u^2 + v^2 \leqq 1, \tag{19}$$

with the following properties:

(i) $x(u, v)$, $y(u, v)$, $z(u, v)$ *form a triple of conjugate harmonic functions in* $u^2+v^2<1$;

(ii) $x(u, v)$, $y(u, v)$, $z(u, v)$ *are continuous in* $u^2+v^2\leqq 1$, *and the equations* (19) *carry* $u^2+v^2=1$ *in a topological way into the Jordan curve* Γ.

By way of explanation, let us recall that a Jordan curve might bound several minimal surfaces, as follows from classical examples. The preceding result expresses a property common to *all* these minimal surfaces.

† See C. Carathéodory, *Conformal Representation*, chapter VI.

Ohio State University,
Columbus, Ohio

SUBHARMONIC FUNCTIONS AND SURFACES OF NEGATIVE CURVATURE*

BY

E. F. BECKENBACH† AND T. RADÓ

Introduction

0.1. Given a piece of surface in general parametric representation

$$x = x(u, v),\ y = y(u, v),\ z = z(u, v),$$

the Gauss curvature K of the surface is given by the familiar formula‡

$$W^4K = \begin{vmatrix} (-\frac{1}{2}G_{uu} + F_{uv} - \frac{1}{2}E_{vv}) & \frac{1}{2}E_u & (F_u - \frac{1}{2}E_v) \\ (F_v - \frac{1}{2}G_u) & E & F \\ \frac{1}{2}G_v & F & G \end{vmatrix}$$

$$- \begin{vmatrix} 0 & \frac{1}{2}E_v & \frac{1}{2}G_u \\ \frac{1}{2}E_v & E & F \\ \frac{1}{2}G_u & F & G \end{vmatrix},$$

where E, F, G are the first fundamental quantities:

$$E = x_u^2 + y_u^2 + z_u^2,\ F = x_ux_v + y_uy_v + z_uz_v,\ G = x_v^2 + y_v^2 + z_v^2,$$

and $W^2 = EG - F^2$. As is usual in differential geometry, we assume throughout this paper that $W \neq 0$ for the representations to be considered.

Suppose now that the surface is given in an isothermic representation; that is to say, suppose that $E = G$, $F = 0$. Put $E = G = \lambda(u, v)$. The assumption $W \neq 0$ is then equivalent to $\lambda(u, v) > 0$. The above formula for K reduces to the form

$$K = \frac{1}{2\lambda^3}(\lambda_u^2 + \lambda_v^2 - \lambda\Delta\lambda),$$

where Δ is the symbol

$$\Delta = \frac{\partial^2}{\partial u^2} + \frac{\partial^2}{\partial v^2}\cdot$$

0.2. By computation it follows that

---

* Presented to the Society, April 14, 1933, under the title *On the isoperimetric inequality*; received by the editors February 16, 1933.

† National Research Fellow.

‡ See W. Blaschke, *Differentialgeometrie*, Berlin, J. Springer, 1930, p. 93.

$$\Delta \log \lambda = \frac{\lambda\Delta\lambda - (\lambda_u^2 + \lambda_v^2)}{\lambda^2}.$$

Hence, we have the formula*

$$K = -\frac{1}{2\lambda}\Delta \log \lambda. \tag{1}$$

Consequently, *if* $K \leqq 0$ *on our surface, then* $\Delta\log\lambda \geqq 0$, *that is to say,* $\log \lambda$ *is subharmonic in the terminology of F. Riesz.*† *Conversely, if* $\log \lambda$ *is subharmonic, then* $K \leqq 0$ *on the surface.*

This relation between subharmonic functions and surfaces of negative curvature suggests geometrical applications of the theory of subharmonic functions. On the other hand, the geometrical interpretation suggests questions concerning subharmonic functions. The purpose of this paper is to present a few results which we have obtained in this way.

0.3. One of our geometrical results is concerned with the isoperimetric inequality. Among all simply-connected plane regions whose boundaries are rectifiable and have a given length l, the circle has the maximum area. This fact may also be stated as follows: if a is the area and l the length of the boundary of a simply-connected plane region, then a and l satisfy the isoperimetric inequality $a \leqq l^2/(4\pi)$. Carleman proved that this same inequality holds for every simply-connected rectifiable piece of a minimal surface.‡ We shall prove that the isoperimetric inequality holds for every simply-connected rectifiable piece of every surface whose Gauss curvature K is $\leqq 0$. This generalization is, in a way, final§; indeed, it is almost trivial (cf. §2.7) that if a surface has the property that every simply-connected piece on it satisfies the isoperimetric inequality, then $K \leqq 0$ on the surface.

We shall make in our work the assumption, customary in differential

* See for instance A. R. Forsyth, *Differential Geometry*, Cambridge University Press, 1912, p. 84.

† See F. Riesz, *Sur les fonctions subharmoniques et leur rapport à la théorie du potentiel* (in two parts), Acta Mathematica, vol. 48 (1926), pp. 329–343, and vol. 54 (1930), pp. 321–360. We shall confine ourselves to the case of continuous subharmonic functions, though Riesz defines them more broadly. For a systematic treatment of the elementary properties of these functions, see P. Montel, *Sur les fonctions convexes et les fonctions sousharmoniques*, Journal de Mathématiques, (9), vol. 7 (1928), pp. 29–60.

‡ T. Carleman, *Zur Theorie der Minimalflächen*, Mathematische Zeitschrift, vol. 9 (1921), pp. 154–160.

§ That is, it is final in so far as the case of surfaces with $K \leqq 0$ is concerned. It is known, however, that for convex regions on a sphere with $K = K_0 > 0$ we have $a \leqq (l^2 + K_0 a^2)/(4\pi)$. See F. Bernstein, *Über die isoperimetrische Eigenschaft des Kreises auf der Kugeloberfläche und in der Ebene*, Mathematische Annalen, vol. 60 (1905), pp. 117–136. There are indications that perhaps the same inequality holds for simply-connected rectifiable regions on surfaces of constant negative curvature, $K = K_0 < 0$. The question arises then as to whether or not for all real K_0 the inequality $a \leqq (l^2 + K_0 a^2)/(4\pi)$ characterizes surfaces with variable curvature $K \leqq K_0$.

geometry, that the surfaces and curves to be considered are analytic. This obviously unnecessary assumption serves the twofold purpose of avoiding certain unessential complications which would obscure the unity and simplicity of the method, and of dodging certain essential difficulties which seem to require a thorough and presumably interesting study.

Besides the isoperimetric inequality, we shall discuss briefly a few theorems which have been first proved for conformal maps of plane regions, have then been extended to conformal maps of minimal surfaces, and will be shown in this paper to hold for conformal maps of surfaces with $K \leqq 0$.

0.4. The following notation will simplify our next statement. If $g(u, v)$ is a continuous function in a domain D, we shall put

$$A(g; u_0, v_0; \rho) = \frac{1}{\pi\rho^2}\iint_{\xi^2+\eta^2<\rho^2} g(u_0 + \xi, v_0 + \eta)d\xi d\eta,$$

$$L(g; u_0, v_0; \rho) = \frac{1}{2\pi}\int_0^{2\pi} g(u_0 + \rho\cos\phi, v_0 + \rho\sin\phi)d\phi,$$

where (u_0, v_0) is the center and ρ is the radius of a circular disc κ: $(u-u_0)^2 + (v-v_0)^2 \leqq \rho^2$ which is comprised in D.

An important inequality, due to Carleman,* can be stated then as follows: if $f(w)$ is an analytic function of $w = u+iv$ in D, then

$$[A(|f|^2; u_0, v_0; \rho)]^{1/2} \leqq L(|f|; u_0, v_0; \rho) \tag{2}$$

for every circular disc comprised in D. We shall show that a function $g(u, v)$, continuous and $\geqq 0$ in D, satisfies the inequality

$$[A(g^2; u_0, v_0; \rho)]^{1/2} \leqq L(g; u_0, v_0; \rho) \tag{3}$$

for every circular disc comprised in D, if and only if $\log g(u, v)$ is subharmonic in the part of D where $g(u, v) > 0$.

We shall use in this paper, as we did in a previous one,† the term function of class PL, meaning a function $g(u, v)$ continuous and $\geqq 0$, and such that $\log g(u, v)$ is subharmonic wherever $g(u, v) > 0$. Then the above inequality (3) expresses a characteristic property of functions of class PL. On account of the formula (1) for K, this analytic fact is then readily seen to be equivalent to the geometric fact that the isoperimetric inequality is characteristic for surfaces with negative curvature (as explained in §0.3).

0.5. It is natural to ask what happens if we replace in (3) the exponents 2 and 1/2 by β and $1/\beta$ respectively, where β is any real number. At the end of §1 we shall make a few very incomplete remarks concerning this question.

* Mathematische Zeitschrift, loc. cit.

† *Subharmonic functions and minimal surfaces*, these Transactions, vol. 35 (1933), pp. 648–661.

1. A CHARACTERIZATION OF FUNCTIONS OF CLASS PL*

1.1. The familiar example of a function of class PL is the absolute value of an analytic function $f(w)$ of the complex variable $w=u+iv$. Indeed, as is well known, $\log |f(w)|$ is a harmonic function of u and v, that is to say, $\Delta \log |f(w)|=0$.

We have the following theorem, due to Carleman.†

If $f(w)$ is continuous in the unit circle $|w|\leqq 1$ and analytic in $|w|<1$, then

$$\iint_{u^2+v^2<1} |f(w)|^2\, dudv \leqq \frac{1}{4\pi}\left[\int_0^{2\pi} |f(e^{i\phi})|\, d\phi\right]^2. \tag{4}$$

The sign of equality in (4) *holds if and only if $f(w)=F'(w)$, where $F(w)$ is a linear function*

$$\frac{aw+b}{cw+d}$$

regular in $|w|\leqq 1$.

1.2. If we write the inequality (4) of Carleman as we did in 0.4, then there arises the following question. Given a domain D in the uv-plane, we ask for all functions $g(u, v)$, which are continuous and $\geqq 0$ in D, and satisfy the inequality (3) for every point (u_0, v_0) in D and for every ρ such that the circular disc $(u-u_0)^2+(v-v_0)^2\leqq\rho^2$ is comprised in D. We shall prove the following

LEMMA. *A function $g(u, v)$, continuous and $\geqq 0$ in a domain D, satisfies the inequality*

$$[A(g^2; u_0, v_0; \rho)]^{1/2} \leqq L(g; u_0, v_0; \rho) \tag{5}$$

for every point (u_0, v_0) in D and for every ρ, such that the circular disc $(u-u_0)^2+(v-v_0)^2\leqq\rho^2$ is comprised in D, if and only if $g(u, v)$ is of class PL in D.

1.3. Let us first prove that if $g(u, v)$ is of class PL in D, then the inequality (5) is satisfied. Suppose first that $g(u, v)>0$ in D. Consider any circular disc κ, comprised in D, with center (u_0, v_0) and radius ρ. Denote by C the perimeter of κ. Put

$$\log g(u, v) = \phi(u, v);$$

then by assumption $\phi(u, v)$ is subharmonic. Let $h(u, v)$ be the harmonic function in κ coinciding with $\phi(u, v)$ on C. Then $\phi(u, v)\leqq h(u, v)$ in κ‡, that is, $g(u, v)\leqq e^{h(u,v)}$ in κ. Consequently,

* For the precise definition and a discussion of elementary facts concerning these functions, see the authors' paper just cited.

† Mathematische Zeitschrift, loc. cit.

‡ This is a general relation between a subharmonic function and a dominating harmonic function. See F. Riesz, Acta Mathematica, loc. cit., first part, p. 331.

$$(6)\qquad [A(g^2; u_0, v_0; \rho)]^{1/2} \leqq [A(e^{2h}; u_0, v_0; \rho)]^{1/2}.$$

Also, $g(u, v) = e^{h(u,v)}$ on C, so that

$$(7)\qquad L(e^h; u_0, v_0; \rho) = L(g; u_0, v_0; \rho).$$

Let $h^*(u, v)$ be the conjugate harmonic function of $h(u, v)$. Then $f(w) = e^{h+ih^*}$ is an analytic function of $w = u + iv$, and $|f(w)| = e^{h(u,v)}$. By Carleman's inequality (2) then

$$(8)\qquad [A(e^{2h}; u_0, v_0; \rho)]^{1/2} \leqq L(e^h; u_0, v_0; \rho).$$

(5) follows from (6), (7), and (8).

Suppose now only that the function $g(u, v)$ of class PL is $\geqq 0$ in D. Consider $g(u, v) + \epsilon$, where ϵ is a constant >0. Then $g(u, v) + \epsilon$ is >0 and of class PL.† Accordingly, the above discussion can be applied to $g(u, v) + \epsilon$, so that (5) holds for this function. As $g(u, v)$ is the uniform limit of $g(u, v) + \epsilon$ as $\epsilon \to 0$, we have (5) for a general $g(u, v)$ of class PL.

1.4. We shall show now that if $g(u, v)$ is a non-negative function defined and continuous in D, and if for every circular disc κ comprised in D, the inequality (5) holds, then $g(u, v)$ is of class PL.

Suppose first that $g(u, v)$ has continuous derivatives of the first and second order, and let these derivatives be denoted by their standard symbols p, q, r, s, t. We assume for convenience that the point (u_0, v_0) under discussion is $(0, 0)$ and denote by p_0, etc., the value of p, etc., at $(0, 0)$. Finally we shall denote by σ_j certain quantities such that $\sigma_j/\rho^2 \to 0$ as $\rho \to 0$, where $\rho^2 = u^2 + v^2$.

We have then, by the finite Taylor expansion,

$$\begin{aligned} g(u, v) &= g_0 + p_0 u + q_0 v + \tfrac{1}{2}(r_0 u^2 + 2 s_0 uv + t_0 v^2) + \sigma_1 \\ &= g_0 + (p_0 \cos\phi + q_0 \sin\phi)\rho \\ &\quad + \tfrac{1}{2}(r_0 \cos^2\phi + 2 s_0 \cos\phi \sin\phi + t_0 \sin^2\phi)\rho^2 + \sigma_1, \\ g(u, v)^2 &= g_0^2 + 2 g_0 (p_0 \cos\phi + q_0 \sin\phi)\rho \\ &\quad + [(p_0^2 + g_0 r_0)\cos^2\phi + 2(p_0 q_0 + g_0 s_0)\cos\phi \sin\phi \\ &\quad + (q_0^2 + g_0 t_0)\sin^2\phi]\rho^2 + \sigma_2, \end{aligned}$$

so that

$$\begin{aligned} L(g; 0, 0; \rho) &= \frac{1}{2\pi}\int_0^{2\pi} g(\rho\cos\phi, \rho\sin\phi)\,d\phi = g_0 + \tfrac{1}{4}\rho^2(r_0 + t_0) + \sigma_3, \\ [L(g; 0, 0; \rho)]^2 &= g_0^2 + \tfrac{1}{2}\rho^2 g_0 (r_0 + t_0) + \sigma_4, \end{aligned}$$

† For the fact that the sum of two functions of class PL is again a function of class PL, see the authors' paper in vol. 35 of these Transactions, pp. 648–661, §1.10.

$$A(g^2; 0, 0; \rho) = \frac{1}{\pi\rho^2}\int_0^\rho \tau d\tau \int_0^{2\pi} g(\tau\cos\phi, \tau\sin\phi)^2 d\phi$$
$$= g_0^2 + \tfrac{1}{4}\rho^2[(p_0^2 + q_0^2) + g_0(r_0 + t_0)] + \sigma_5.$$

By assumption, then,

$$g_0^2 + \tfrac{1}{4}\rho^2[(p_0^2 + q_0^2) + g_0(r_0 + t_0)] + \sigma_5 \leqq g_0^2 + \tfrac{1}{2}\rho^2 g_0(r_0 + t_0) + \sigma_4,$$

or

$$(p_0^2 + q_0^2) - g_0(r_0 + t_0) \leqq 4(\sigma_4 - \sigma_5)/\rho^2.$$

The right-hand member of this last inequality $\to 0$ as $\rho \to 0$, so that the left-hand member is $\leqq 0$. Since any point of D can be taken as (u_0, v_0), we have then

$$g(r + t) - (p^2 + q^2) \geqq 0$$

in D. Hence $g(u, v)$ is of class PL, since by computation (cf. §0.2)

$$\Delta \log g = \frac{g(r + t) - (p^2 + q^2)}{g^2},$$

wherever $g > 0$.

1.5. Suppose now† that $g(u, v)$ has continuous derivatives of only the first order, but otherwise satisfies the conditions of §1.4. For a small fixed $\tau > 0$, put

$$g(u, v; \tau) = \frac{1}{\pi\tau^2}\iint_{\xi^2+\eta^2<\tau^2} g(u + \xi, v + \eta) d\xi d\eta.$$

(Of course $g(u, v; \tau)$ can be defined thus for only a subdomain D' of D, but this is of no consequence since τ is arbitrarily small.) That this function $g(u, v; \tau)$ also satisfies (5) follows from Minkowski's inequality.‡ Furthermore $g(u, v; \tau)$ has continuous derivatives of the second order.§ Hence $g(u, v; \tau)$

† The assumptions of §1.4 are sufficient for the applications to differential geometry which we shall make in §2, so that the reader interested primarily in those applications can omit §1.5 and §1.6 without loss of continuity in the discussion.

‡ The necessary inequality follows, by a familiar passage to the limit, from the inequality

$$\left(\sum_{k=1}^{n}\left(\sum_{j=1}^{m} a_{kj}\right)^2\right)^{1/2} \leqq \sum_{j=1}^{m}\left(\sum_{k=1}^{n} a_{kj}^2\right)^{1/2}$$

which has the geometrical significance that the length of a polygonal line is at least as great as that of the line segment joining its end points.

§ Concerning the properties and applications of this approximation by integral means, see E. Levi, *Sopra una proprietà caratteristica delle funzione armoniche*, Atti della Reale Accademia dei Lincei, vol. 18 (1909), pp. 10–15; H. E. Bray, *Proof of a formula for an area*, Bulletin of the American Mathematical Society, vol. 29 (1923), pp. 264–270; T. Radó, *Remarque sur les fonctions subharmoniques*, Paris Comptes Rendus, vol. 186, pp. 346–348; T. Radó, *Sur le calcul de l'aire des surfaces courbes*, Fundamenta Mathematicae, vol. 10 (1927), pp. 197–210; F. Riesz, loc. cit., second part, pp. 342–345.

satisfies all the conditions of §1.4 and so is of class PL. Since $g(u, v; \tau) \rightrightarrows g(u, v)$ as $\tau \to 0$ it follows that $g(u, v)$ is of class PL.

1.6. Suppose finally that $g(u, v)$ is only continuous, but otherwise satisfies the conditions of §1.4. Then $g(u, v; \tau)$, defined as above, has continuous first derivatives†, and hence it satisfies the assumptions of §1.5. According to §1.5, $g(u, v; \tau)$ is of class PL and consequently its uniform limit $g(u, v)$ is of class PL.

1.7. With regard to an application which we shall make in §2, we need a slight (and incomplete) discussion of the sign of equality in (5). Suppose that $g(u, v)$ is continuous and positive in $(u-u_0)^2+(v-v_0)^2 \leqq \rho^2$ and that $g(u, v)$ is of class PL in $(u-u_0)^2+(v-v_0)^2 < \rho^2$. Suppose that

$$[A(g^2; u_0, v_0; \rho)]^{1/2} = L(g; u_0, v_0; \rho). \tag{9}$$

Then $g(u, v) \equiv |F'(w)|$, where $F(w)$ is a linear function

$$\frac{aw + b}{cw + d}$$

which is regular in $|w - w_0| \leqq \rho$ and which does not reduce to a constant.

Indeed, if we go through the discussion in §1.3, we find that in order to have (9), we must have (with the notations of §1.3)

$$g(u, v) \equiv | f(w) |,$$

where $f(w)$ satisfies the inequality (4) of Carleman with the sign of equality. On account of the theorem of Carleman, we have then $f(w) = F'(w)$, where $F(w)$ has the desired form. This $F(w)$ cannot reduce to a constant at present, since then it would follow that $g(u, v) \equiv |F'(w)| = 0$, while we supposed that $g(u, v) > 0$ throughout.

1.8. The question arises as to the significance of the inequality (5) if we replace the exponent 2 by a general (real) exponent β. The case $\beta = 1$ can be settled easily; the reasoning used above in the case $\beta = 2$ applies directly.‡ There follows

A function $g(u, v)$, continuous in a domain D, satisfies there the inequality

$$A(g; u_0, v_0; \rho) \leqq L(g; u_0, v_0; \rho)$$

for every point (u_0, v_0) in D and for every ρ, such that the circular disc $(u-u_0)^2 + (v-v_0)^2 \leqq \rho^2$ is comprised in D, if and only if $g(u, v)$ is subharmonic in D.

1.9. For values of β other than 1 and 2, the method of §1.4, §1.5, §1.6 yields theorems whose statements vary according to the location of β with

† See third footnote on p. 667.

‡ Actually, though, there is a much simpler way of handling the case $\beta = 1$.

respect to the special values 0, 1, 2. By way of illustration, we mention the following statements.

Suppose $g(u, v)$ is continuous and $\geqq 0$ in a domain D. Suppose that for a certain exponent β the inequality

$$[A(g^\beta; u_0, v_0; \rho)]^{1/\beta} \leqq L(g; u_0, v_0; \rho) \tag{10}$$

holds for every point (u_0, v_0) in D and for every ρ such that the circular disc $(u-u_0)^2+(v-v_0)^2 \leqq \rho^2$ is comprised in D.

If $1<\beta<2$, *then it follows that* $g^{2-\beta}$ *is subharmonic in* D.

If $\beta>2$, *it follows that* $1/g^{\beta-2}$ *is superharmonic.*

In a general way, the greater β, the stronger the inference will be as to the subharmonic character of $g(u, v)$. For $\beta<1$, $g(u, v)$ need not be subharmonic.

For $\beta=1$ and $\beta=2$ the inequality (10) has been shown, in what precedes, to be a necessary and sufficient criterion for a certain subharmonic property. An equally complete discussion for a general exponent might lead to interesting questions.

2. Applications to surfaces of negative curvature

2.1. Let there be given a piece of surface S in a representation

$$S\colon x = x(u, v),\ y = y(u, v),\ z = z(u, v),\ u^2 + v^2 \leqq \rho^2, \tag{11}$$

with the following properties.

(a) $x(u, v), y(u, v), z(u, v)$ and their first partial derivatives are continuous in $u^2+v^2 \leqq \rho^2$.

(b) In $u^2+v^2<\rho^2$, $x(u, v)$, $y(u, v)$, $z(u, v)$ have continuous partial derivatives of the third order.

(c) The representation (11) is isothermic, that is to say, $E=G$, $F=0$, in $u^2+v^2 \leqq \rho^2$. We put $E=G=\lambda(u, v)$. Then $\lambda \geqq 0$; but we suppose that $\lambda>0$ in $u^2+v^2 \leqq \rho^2$.

2.2. Lemma. *If the Gauss curvature K of the surface S, given in a representation as described in* §2.1, *is* $\leqq 0$, *then the area a and the perimeter l of S satisfy the inequality* $a \leqq l^2/(4\pi)$. *The sign of equality holds if and only if* $K\equiv 0$, *and S is a geodesic circle (that is to say, S is a developable and there exists a point O on S such that the geodesic distance of O from every point of the perimeter of S is the same).*

The proof is as follows. With the notation of §0.4 we have

$$a = \pi\rho^2 A(\lambda; 0, 0; \rho),\ l = 2\pi\rho L(\lambda^{1/2}; 0, 0; \rho). \tag{12}$$

On account of the assumption $K \leqq 0$, the function $\lambda(u, v)$ is of class PL (see

§0.2 and the definition of functions of class PL). Hence† the function $\lambda(u, v)^{1/2}$ is also of class PL. From §1.2 it follows, therefore, for $g=\lambda^{1/2}$, that

$$[A(\lambda; 0, 0; \rho)]^{1/2} \leqq L(\lambda^{1/2}; 0, 0; \rho). \tag{13}$$

The inequality $a \leqq l^2/(4\pi)$ follows now immediately from (12) and (13).

Suppose now that we have $a=l^2/(4\pi)$. Then we must have the sign of equality in (13). Consequently (see §1.7) we have

$$\lambda(u, v)^{1/2} = |F'(w)|, \tag{14}$$

where $F(w)$ has the form $(aw+b)/(cw+d)$, and $F(w)$ is regular and not constant in $|w| \leqq \rho$. Hence the equation $w^*=F(w)$ carries the circle $|w| \leqq \rho$ in a one-to-one and conformal way into a certain circular disc κ^* in the $w^*=u^*+iv^*$ plane. Introducing u^*, v^* as new parameters, we obtain the equations of S in the form

$$S:\ x = \xi(u^*, v^*),\ y = \eta(u^*, v^*),\ z = \zeta(u^*, v^*),\ (u^*, v^*) \text{ in } \kappa^*. \tag{15}$$

Since we passed from the isothermic parameters u, v to the new parameters u^*, v^* by a conformal map, it follows that u^*, v^* are also isothermic parameters. Hence if we denote by E^*, F^*, G^* the first fundamental quantities relative to the representation (15), we have $E^*=G^*$, $F^*=0$. If we put $E^*=G^*=\lambda^*(u^*, v^*)$, then we have, by simple computation,

$$\lambda^{*1/2} = \lambda^{1/2}\left|\frac{dw}{dw^*}\right| = \lambda^{1/2}/|F'(w)| = 1,$$

on account of (14). Hence $E^*=G^*=1$, $F^*=0$. That is to say, the representation (15) is an isometric map of S (every arc on S has the same length as its image).

2.3. In order to apply the lemma of §2.2 to a given piece of surface, we have to represent the surface as required in §2.1. Thus it is necessary to refer to existence theorems on conformal mapping, and the validity of the isoperimetric inequality $a \leqq l^2/(4\pi)$ is made to depend upon the available results concerning the theory of conformal mapping. Since we are unable at this time to prove the most general statement which is likely to be true, we restrict ourselves to the following theorem which might be considered as perfectly general according to the usual standards in differential geometry.

2.4. THEOREM. *Let there be given an analytic surface in the xyz-space, that is to say, a surface which admits, in the vicinity of every one of its points, a representation* $x=x(u, v)$, $y=y(u, v)$, $z=z(u, v)$, *where* $x(u, v)$, $y(u, v)$, $z(u, v)$

† Every positive power of a function of class PL is again a function of class PL; see the authors' paper in these Transactions, vol. 35, pp. 648–661, §1.8.

are analytic functions of u, v, and where $EG-F^2>0$. Denote by K the Gauss curvature of the surface. Then $K\leqq 0$ is a necessary and sufficient condition that the area a and the perimeter l of every simply-connected portion, bounded by an analytic curve, of the surface satisfy the isoperimetric inequality $a\leqq l^2/(4\pi)$.

2.5. To prove this theorem, suppose first that $K\leqq 0$. Let S be any simply-connected portion of the surface, bounded by an analytic curve. Take a simply-connected open portion S^* of the surface, such that S is interior to S^*. On account of general theorems, S^* admits of an isothermic representation

$$S^*:\ x=\xi(u^*,v^*),\ y=\eta(u^*,v^*),\ z=\zeta(u^*,v^*),\ u^{*2}+v^{*2}<1,$$

where $E^*=G^*>0$, $F^*=0$, and ξ, η, ζ are analytic functions of u^*, v^*. The portion S appears in this map as a Jordan region R^* in $u^{*2}+v^{*2}<1$, bounded by an analytic Jordan curve C^*. We map then R^* in a one-to-one and conformal way upon $u^2+v^2\leqq 1$; on account of the analyticity of C^*, this map remains analytic on $u^2+v^2=1$. Thus we obtain a map of S as required in §2.1, and then the lemma of §2.2 gives the desired inequality $a\leqq l^2/(4\pi)$.

2.6.† Suppose, conversely, that we have $a\leqq l^2/(4\pi)$ for every portion S of a surface as described in §2.4. Take such a portion S. Applying the construction of §2.5, we obtain for S a representation

$$S:\ x=x(u,v),\ y=y(u,v),\ z=z(u,v),\ u^2+v^2\leqq 1, \tag{16}$$

with the properties required in §2.1. If $\kappa:(u-u_0)^2+(v-v_0)^2\leqq\rho^2$ is any circular disc comprised in $u^2+v^2<1$, then there corresponds to κ, by means of (16), a portion S_0 whose area a_0 and perimeter l_0 satisfy by assumption the inequality $a_0\leqq l_0^2/(4\pi)$. If we use again the notation $E=G=\lambda(u,v)$, then

$$a_0=\pi\rho^2A(\lambda;\,u_0,v_0;\,\rho),\ l_0=2\pi\rho L(\lambda^{1/2};\,u_0,v_0;\,\rho),$$

and hence $a_0\leqq l_0^2/(4\pi)$ implies that

$$[A(\lambda;\,u_0,v_0;\,\rho)]^{1/2}\leqq L(\lambda^{1/2};\,u_0,v_0;\,\rho).$$

Since this holds for every circular disc $(u-u_0)^2+(v-v_0)^2\leqq\rho^2$ comprised in $u^2+v^2<1$, it follows (see §1.2) that $\lambda^{1/2}$ and consequently‡ λ is of class PL in $u^2+v^2<1$. Hence (see §0.2) $K\leqq 0$ on S. Since S was any portion of the given surface, this proves that $K\leqq 0$ on the whole surface.

2.7. The reasoning of §2.6 can be replaced by the following argument. Take any point O on the surface and denote by S_ρ the portion of the surface which consists of the points of the surface within and on the geodesic circle

† A differential geometer will probably find the proof of §2.7 preferable.

‡ See footnote on p. 670.

with center O and radius ρ. Then the area and perimeter of S_ρ are functions of ρ which admit of developments beginning as follows†:

$$a(\rho) = \pi\rho^2 - \frac{1}{12}\pi K_0\rho^4 + \cdots,$$

$$l(\rho) = 2\pi\rho - \tfrac{1}{3}\pi K_0\rho^3 + \cdots,$$

where K_0 is the Gauss curvature at the point O. We have then

$$a(\rho) - \frac{1}{4\pi}l(\rho)^2 = \frac{1}{4}\pi K_0\rho^4 + \cdots,$$

or

$$K_0 = \frac{4}{\pi}\lim_{\rho\to 0}\frac{a(\rho) - \frac{1}{4\pi}l(\rho)^2}{\rho^4}\cdot$$

Since, by assumption, the numerator is $\leqq 0$, this proves that $K_0 \leqq 0$. Since O is any point on the surface, we have then $K \leqq 0$ on the whole surface.

2.8. Let us now consider the sign of equality in the isoperimetric inequality. In order to illustrate a very trivial point, let us consider a Jordan region in the plane, bounded by a rectifiable curve which is not a circle. Then we have $a < l^2/(4\pi)$. Putting some hills on this plane region, we can keep the perimeter l fixed and increase the area until we have $a = l^2/(4\pi)$. Since our hills were otherwise quite arbitrary, it is clear that from $a = l^2/(4\pi)$ alone we cannot conclude anything concerning the surface. On the other hand, if we restrict ourselves to analytic surfaces with $K \leqq 0$, and if we use the fact that $K \equiv 0$ on such a surface as soon as $K \equiv 0$ on any subregion, then the lemma of §2.2 yields immediately the following result.

If an analytic surface with $K \leqq 0$ contains some portion for which the sign of equality holds in the isoperimetric inequality $a \leqq l^2/(4\pi)$, then $K \equiv 0$ on the surface, and $a = l^2/(4\pi)$ holds only for the geodesic circles.

2.9. In what precedes, we extended a theorem, previously proved only for minimal surfaces, to surfaces with $K \leqq 0$. A systematic study of similar generalizations might lead to interesting results. We mention here a few immediate facts.

Let S be a piece of surface with $K \leqq 0$, which admits an isothermic representation

$$(17) \qquad S\colon x = x(u, v),\ y = y(u, v),\ z = z(u, v),\ u^2 + v^2 \leqq \rho^2,$$

with the properties described in §2.1. Put again $E = G = \lambda(u, v)$, and suppose

† See for instance L. P. Eisenhart, ***Differential Geometry***, Ginn and Company, 1909, p. 209.

that $\lambda(0, 0)=1$ (that is to say, that the linear magnification is unity at the origin). Denote by $l(r)$ the length of the image of $u^2+v^2=r^2$, and by $a(r)$ the area of the image of $u^2+v^2<r^2$. Then

(a) *$l(r)$ is an increasing function of r*†;

(b) $l(r)\geqq 2\pi r$‡;

(c) $a(r)\geqq \pi r^2$.‡

We have

$$l(r) = \int_0^{2\pi} \lambda(r\cos\phi, r\sin\phi)^{1/2} r\,d\phi.$$

Since $K\leqq 0$, it follows§ that $\lambda(u, v)^{1/2}$ is of class PL. Also (see §1.1), $r=|u+iv|$ is of class PL. Therefore‖ $r\lambda^{1/2}$ is of class PL, and consequently¶ $r\lambda^{1/2}$ is subharmonic. (a) follows then from the above expression for $l(r)$ and from the fact that the integral mean of a subharmonic function is an increasing function of r.††

To prove (b) and (c), observe that

$$\begin{aligned} a(r) &= \pi r^2 A(\lambda; 0, 0; r), \\ l(r) &= 2\pi r L(\lambda^{1/2}; 0, 0; r). \end{aligned} \tag{18}$$

On account of $K\leqq 0$, λ and consequently $\lambda^{1/2}$ are subharmonic. Hence‡‡

$$\begin{aligned} 1 &= \lambda(0, 0) \leqq A(\lambda; 0, 0; r), \\ 1 &= \lambda(0, 0)^{1/2} \leqq L(\lambda^{1/2}; 0, 0; r). \end{aligned} \tag{19}$$

Thus (b) and (c) follow from (18) and (19).

---

† For the plane case, see L. Bieberbach, *Über die konforme Kreisabbildung nahezu kreisförmiger Bereiche*, Berlin Sitzungsberichte, 1924, pp. 181–188; for minimal surfaces, see T. Radó, *Some remarks on the problem of Plateau*, Proceedings of the National Academy of Sciences, vol. 16 (1930), pp. 242–248, and *On Plateau's problem*, Annals of Mathematics, vol. 31 (1930), pp. 457–469.

‡ For the plane case, see L. Bieberbach, Palermo Rendiconti, vol. 38 (1914), pp. 98–112; for minimal surfaces, see E. F. Beckenbach, *The area and boundary of minimal surfaces*, Annals of Mathematics, vol. 33 (1932), pp. 658–664.

§ See the authors' paper in these Transactions, vol. 35, pp. 648–661, §1.8.

‖ The product of two functions of class PL is again a function of class PL; see the authors' paper in these Transactions, vol. 35, pp. 648–661, §1.8.

¶ Every positive power of a function of class PL is a subharmonic function; see the authors' paper in these Transactions, vol. 35, pp. 648–661, §§1.7 and 1.8.

†† See F. Riesz, Acta Mathematica, loc. cit., first part, p. 338.

‡‡ These inequalities express two of several clearly equivalent definitions of subharmonic functions. See J. E. Littlewood, *On the definition of a subharmonic function*, London Mathematical Society Journal, vol. 2 (1927), pp. 189–192.

Corollary. *If the sign of equality holds in* (b) *or* (c) *for any value of* r, $0<r\leqq\rho$, *then* $\lambda(u, v)\equiv 1$, *so that* (see §2.2) *the map* (17) *is isometric. In other words, S is a developable piece of surface and is a geodesic circle given in isometric representation.*

To see this, consider for instance the sign of equality in (c); then

$$\lambda(0, 0) = \frac{1}{\pi r^2}\iint_{u^2+v^2<r^2} \lambda(u, v)\,du\,dv.$$

Consequently $\lambda(0, 0)=h(0, 0)$, where $h(u, v)$ is the harmonic function in $u^2+v^2<r^2$ coinciding with $\lambda(u, v)$ on $u^2+v^2=r^2$, and therefore† $\lambda(u, v)\equiv h(u, v)$, so that

$$\Delta\lambda(u, v) = 0.$$

But $\lambda(u, v)$ is also of class PL, so that (see §0.2)

$$\lambda\Delta\lambda - (\lambda_u^2 + \lambda_v^2) \geqq 0.$$

Consequently $\lambda_u{}^2+\lambda_v{}^2\leqq 0$ and therefore $\lambda(u, v)$ is constant. But $\lambda(0, 0)=1$, so that $\lambda(u, v)\equiv 1$. The same argument holds for the sign of equality in (b), with $\lambda(u, v)^{1/2}$ in place of $\lambda(u, v)$.

† $\lambda(u, v)$ is subharmonic, and therefore, by the definition of subharmonic functions (see F. Riesz, Acta Mathematica, loc. cit., first part, p. 331) $\psi(u, v)=\lambda(u, v)-h(u, v)$ is subharmonic. We have $\psi(r\cos\phi, r\sin\phi)\equiv 0$, $\psi(0, 0)=0$. But a subharmonic function cannot attain an interior maximum unless it is identically constant (see above reference to F. Riesz, p. 331). Therefore $\psi(u, v)\equiv 0$, $\lambda(u, v)\equiv h(u, v)$.

Ohio State University,
Columbus, Ohio

A TRANSFORMATION OF THE PROBLEM OF LAGRANGE IN THE CALCULUS OF VARIATIONS*

BY

LAWRENCE M. GRAVES

By means of a simple transformation suggested by Bliss, the problem of Lagrange may be reduced to one in which the side conditions are integral equations rather than differential equations, and no derivatives enter explicitly. A multiplier rule for the transformed problem is derived below, in which the multipliers are all constants. When the inverse transformation is applied to this multiplier rule, formulas are obtained for the non-constant multipliers occurring in the ordinary form of the Lagrange multiplier rule, and it is seen that the constant multipliers obtained here may be identified with certain constants appearing in the ordinary form of the rule.

In connection with his applications of the calculus of variations to problems in economics, Roos† has been led to consider a generalization of the problem of Lagrange in which integro-differential equations occur among the side conditions. The transformation and analysis given below apply with equal facility to Roos' problem.

For normal arcs an analogue of the Weierstrass condition is derived for the transformed problem. It is not necessary to assume that the minimizing arc is normal on sub-arcs. For such problems as that of Roos, no generalization of the Jacobi-Mayer condition has to my knowledge yet been obtained, though several attempts have been made.

1. The transformation of the problem. We shall start with the problem suggested by Roos, in the following form: To find necessary conditions on a curve

$$y_i = y_i(x) \qquad (i = 1, \cdots, n;\ x_1 \leqq x \leqq x_2),$$

which minimizes an integral

$$I = \int_{x_1}^{x_2} f(x, y, y')dx$$

in a certain class of curves satisfying the integro-differential equations

* Presented to the Society, December 31, 1930; received by the editors November 21, 1932.

† *Generalized Lagrange problems in the calculus of variations*, these Transactions, vol. 30 (1928), pp. 360–384.

$$\phi_\alpha[x, y(x), y'(x), u(x)] = 0 \qquad (\alpha = 1, \cdots, m < n),$$

$$u_\gamma(x) = \int_{x_1}^{x} P_\gamma[x, s, y(s), y'(s)]ds \qquad (\gamma = 1, \cdots, q),$$

and the end conditions $y_i(x_1) = y_{i1}$, $y_i(x_2) = y_{i2}$. The functions f, ϕ_α, and P_γ are supposed to have continuous first partial derivatives with respect to all their arguments in a certain region R of $(2n+q+2)$-dimensional space. The curves admitted to consideration are supposed to be of class D', i.e., the functions $y_i(x)$ are continuous and their derivatives $y_i'(x)$ have at most a finite number of ordinary finite discontinuities. Admissible curves are also supposed to have all their elements

$$[x, s, y(s), y'(s), u(s)] \qquad (x_1 \leqq s \leqq x \leqq x_2)$$

interior to the region R. We shall suppose also that along the minimizing curve the matrix of partial derivatives $\phi_{\alpha y_i'}$ $(\alpha = 1, \cdots, m; i = 1, \cdots, n)$ has rank m. For simplicity we suppose that the minimizing curve itself is of class C'.

Then as Bliss* has shown, additional functions $\phi_r(x, y')$ $(r = m+1, \cdots, n)$ may be adjoined, with the same continuity properties as the original functions ϕ_α, so that the functional determinant $|\phi_{iy_j'}|$ does not vanish along the minimizing curve. Hence the equations $\phi_i(x, y, y', u) = z_i$ have a unique continuous solution

$$y_i' = \psi_i(x, y, u, z) \tag{1}$$

with (x, y, y', u, z) near the values along the minimizing curve, and the functions ψ_i have continuous first partial derivatives. If equations (1) are used to eliminate the y_i', the integral I becomes

$$I = \int_{x_1}^{x_2} g[s, y(s), u(s), z(s)]ds$$

and the side conditions become

$$y_i(x) = y_{i1} + \int_{x_1}^{x} \psi_i[s, y(s), u(s), z(s)]ds \quad (i = 1, \cdots, n),$$

$$u_\gamma(x) = \int_{x_1}^{x} Q_\gamma[x, s, y(s), u(s), z(s)]ds \qquad (\gamma = 1, \cdots, q),$$

$$z_\alpha(x) = 0 \qquad (\alpha = 1, \cdots, m).$$

The end conditions are $y_i(x_2) = y_{i2}$ $(i = 1, \cdots, n)$.

---

* *The problem of Mayer with variable end points*, these Transactions, vol. 19 (1918), p. 312.

2. **The multiplier rule, and an analogue of the Weierstrass condition, for the transformed problem.** We shall now consider the new form of the problem on its own merits, and in order to simplify the notation in this section, we reformulate it as follows: To find necessary conditions on a "curve"

$$y_i = y_i(x), \qquad z_r = z_r(x) \qquad (i = 1, \cdots, n; r = 1, \cdots, \nu), \tag{2}$$

which minimizes an integral

$$I = \int_{x_1}^{x_2} g[s, y(s), z(s)]ds$$

in a certain class of curves satisfying the integral equations

$$y_i(x) = y_{i1} + \int_{x_1}^{x} \psi_i[x, s, y(s), z(s)]ds \qquad (i = 1, \cdots, n; x_1 \leqq x \leqq x_2) \tag{3}$$

and the end conditions

$$y_i(x_2) = y_{i2} \qquad (i = 1, \cdots, p \leqq n). \tag{4}$$

The functions $g(s, y, z)$ and $\psi_i(x, s, y, z)$ are supposed to be continuous and to have continuous partial derivatives with respect to their arguments y_i and z_r in a certain region R of (x, s, y, z) space. The curves (2) admitted to consideration are supposed to have all their elements $(x, s, y(s), z(s))$ interior to R, and the functions $y_i(x)$ are supposed to be continuous, while the functions $z_r(x)$ have at most a finite number of ordinary finite discontinuities.

Under these circumstances, the equations (3) have a unique solution $y_i(x) = Y_i[z|x]$ for each set of functions $z_r(x)$ near those associated with the minimizing curve, and the functionals Y_i have differentials* $\eta_i(x) = dY_i[z; \zeta|x]$ which satisfy the equations of variation

$$\eta_i(x) = \int_{x_1}^{x} \psi_{iy_j}(x, s)\eta_j(s)ds + \int_{x_1}^{x} \psi_{iz_r}(x, s)\zeta_r(s)ds. \tag{5}$$

Here and elsewhere we abbreviate such expressions as $\psi_{iy_j}[x, s, y(s), z(s)]$ to $\psi_{iy_j}(x, s)$. When the functionals $Y_i[z]$ are substituted in the integral I, it becomes a functional $J[z]$, which is to be minimized in the class of functions $z_r(x)$ for which $Y_i[z|x_2] = y_{i2}$ $(i = 1, \cdots, p)$. The functional $J[z]$ also has a differential given by

$$dJ[z; \zeta] = \int_{x_1}^{x_2} \{g_{y_j}(s)\eta_j(s) + g_{z_r}(s)\zeta_r(s)\}ds,$$

* See, e.g., Graves, *Implicit functions and differential equations in general analysis*, these Transactions, vol. 29 (1927), pp. 514–552.

where the η_j are determined by equations (5). If $J[z]$ is a minimum, the usual argument shows that there exist constants $l_0, c_1, \cdots, c_p$, not all zero, such that

$$l_0 dJ[z;\zeta] + \sum_{i=1}^{p} c_i dY_i[z;\zeta \mid x_2] = 0 \tag{6}$$

for all functions ζ_r having only a finite number of finite discontinuities. If we set

$$G(s, y, z) = l_0 g(s, y, z) + \sum_{i=1}^{p} c_i \psi_i(x_2, s, y, z), \tag{7}$$

equation (6) becomes

$$\int_{x_1}^{x_2} \{G_{y_j}(s)\eta_j(s) + G_{z_r}(s)\zeta_r(s)\} ds = 0. \tag{8}$$

Let $S_{ij}(x, s)$ denote the reciprocal kernel matrix for the Volterra system (5), so that its solution is given by

$$\begin{aligned} \eta_i(x) = \int_{x_1}^{x} &\psi_{iz_r}(x, s)\zeta_r(s) ds \\ &- \int_{x_1}^{x} S_{ij}(x, t) \int_{x_1}^{t} \psi_{jz_r}(t, s)\zeta_r(s) ds dt. \end{aligned} \tag{9}$$

By substituting (9) in (8) and making certain interchanges in the order of integration, we find

$$\int_{x_1}^{x_2} \zeta_r(x)\lambda_r(x) dx = 0$$

for all $\zeta_r(x)$, where

$$\begin{aligned} \lambda_r(x) = G_{z_r}(x) + \int_{x}^{x_2} &G_{y_j}(t)\psi_{jz_r}(t, x) dt \\ &- \int_{x}^{x_2} \int_{x}^{u} G_{y_j}(u) S_{jk}(u, t)\psi_{kz_r}(t, x) dt du. \end{aligned} \tag{10}$$

Hence we have proved the

ANALOGUE OF THE LAGRANGE MULTIPLIER RULE. *If the functions $y_i(x)$, $z_r(x)$ minimize the integral I in the class of all such functions satisfying the integral equations* (3) *and the end conditions* (4), *then there exist constants l_0, $c_1, \cdots, c_p$, not all zero, such that*

$$
\begin{aligned}
&G_{z_r}(x) + \int_x^{x_2} G_{y_j}(t)\psi_{jz_r}(t, x)dt \\
(11)\qquad &- \int_x^{x_2}\int_x^{u} G_{y_j}(u)S_{jk}(u, t)\psi_{kz_r}(t, x)dtdu = 0
\end{aligned}
$$

$$(x_1 \leqq x \leqq x_2;\ r = 1, \cdots, \nu),$$

where $G(s, y, z)$ is defined by equation (7), *and $S_{jk}(x, s)$ is the reciprocal kernel matrix for the system* (5).

We shall say that a curve $y_i = y_i(x)$, $z_r = z_r(x)$ $(x_1 \leqq x \leqq x_2)$ is *normal* in case there exist p sets of variations $\eta_{i\sigma}(x)$, $\zeta_{r\sigma}(x)$ $(\sigma = 1, \cdots, p)$, satisfying the equations of variation (5) and such that the determinant $|\eta_{i\sigma}(x_2)|$ $(i, \sigma = 1, \cdots, p)$ does not vanish. The usual considerations show that an arc is normal if and only if it has no set of multipliers $l_0, c_1, \cdots, c_p$, with $l_0 = 0$, with which it satisfies the equations (11). For a normal minimizing arc we may always assume $l_0 = 1$, and then the remaining multipliers are uniquely determined.

ANALOGUE OF THE WEIERSTRASS CONDITION. *If the minimizing curve for our problem is normal, and if l_0 is taken equal to unity, then for every element (x, y, z) of the minimizing curve and for arbitrary numbers Z_r, the expression*

$$
\begin{aligned}
&G(x, y, Z) - G(x, y, z) \\
&\quad + \int_x^{x_2}\left[G_{y_j}(t) - \int_t^{x_2} G_{y_i}(s)S_{ij}(s, t)ds\right][\psi_j(t, x, y, Z) - \psi_j(t, x, y, z)]dt
\end{aligned}
$$

cannot be negative.

This theorem may be proved by the method of the author's paper *The Weierstrass condition for the problem of Bolza in the calculus of variations*†, as follows. Let $\eta_{j\sigma}$, $\zeta_{r\sigma}$ $(\sigma = 1, \cdots, p)$ be an admissible set of functions satisfying the equations of variation (5), and such that the determinant $|\eta_{i\sigma}(x_2)|$ $\neq 0$, where $i, \sigma = 1, \cdots, p$. Let $x_1 < x_3 < x_2$, and

$$
\begin{aligned}
z_r^*(x, \beta, \epsilon) &= z_r(x) + \epsilon_\sigma\zeta_{r\sigma}(x) &&\text{on } x_1 \leqq x \leqq x_3,\ x_3 + \beta < x \leqq x_2, \\
&= Z_r &&\text{on } x_3 < x \leqq x_3 + \beta.
\end{aligned}
$$

When the functions z^* are substituted for z in equations (3) these equations determine functions $y_i = y_i^*(x, \beta, \epsilon)$ defined for $x_1 \leqq x \leqq x_2$, (β, ϵ) near $(0, 0)$, which are continuous and have partial derivatives with respect to β and ϵ_σ which are continuous except that the partial derivatives $y_{i\beta}^*$ may be discon-

† Annals of Mathematics, vol. 33 (1932), pp. 747–752.

tinuous in x at $x=x_3+\beta$. Set $I(y^*, z^*)=I(\beta, \epsilon)$. We are supposing that $I(\beta, \epsilon)$ has a minimum for $\beta=\epsilon_\sigma=0$. Then if the equations

$$I(\beta, \epsilon) = I(0, 0) + v, \quad y_i^*(x_2, \beta, \epsilon) = y_{i2} \qquad (i = 1, \cdots, p) \tag{12}$$

have a solution $\beta(v)$, $\epsilon_\sigma(v)$ near $v=0$, we must have $\beta'(0)\geqq 0$. By differentiating equations (12) with respect to v, we find for $\beta=\epsilon_\sigma=0$,

$$\begin{aligned} I_\beta\beta' + I_{\epsilon_\sigma}\epsilon_\sigma' &= 1, \\ \theta_i(x_2)\beta' + \eta_{i\sigma}(x_2)\epsilon_\sigma' &= 0 \qquad (i = 1, \cdots, p), \end{aligned}$$

where $\theta_i(x)\equiv y_{i\beta}^*(x, 0, 0)$. Multiply the last p equations by the constants $c_1, \cdots, c_p$ respectively, and add to the first. By equation (6) the result is

$$\beta'\left[I_\beta+\sum_{i=1}^{p}c_i\theta_i(x_2)\right]=1.$$

Hence

$$E \equiv I_\beta + \sum_{i=1}^{p}c_i\theta_i(x_2) \geqq 0.$$

Now the functions $\theta_i(x)$ satisfy the equations

$$\begin{aligned} \theta_i(x) &= 0 \qquad (x_1 \leqq x < x_3), \\ \theta_i(x) &= \Delta_i(x) + \int_{x_3}^{x} \psi_{iy_j}(x, s)\theta_j(s)ds \qquad (x_3 \leqq x \leqq x_2), \end{aligned}$$

where $\Delta_i(x)=\psi_i(x, x_3, y_3, Z)-\psi_i(x, x_3, y_3, z_3)$, $y_3=y(x_3)$, $z_3=z(x_3)$. Hence by use of the reciprocal kernel S_{ij} of ψ_{iy_j},

$$\theta_i(x) = \Delta_i(x) - \int_{x_3}^{x} S_{ij}(x, t)\Delta_j(t)dt.$$

By direct calculation

$$I_\beta = g(x_3, y_3, Z) - g(x_3, y_3, z_3) + \int_{x_3}^{x_2} g_{y_j}(s)\theta_j(s)ds.$$

Combining these results we find

$$\begin{aligned} E = {} & G(x_3, y_3, Z) - G(x_3, y_3, z_3) \\ & + \int_{x_3}^{x_2} G_{y_j}(x)\left[\Delta_j(x) - \int_{x_3}^{x} S_{jk}(x, t)\Delta_k(t)dt\right]dx, \end{aligned}$$

which reduces by an interchange of order of integration to the expression given in the theorem.

3. **Application of the inverse transformation to the new multiplier rule.** Returning now to the problem of §1, we shall for simplicity consider only the case when the functions ϕ_α are independent of the u's, that is to say, the ordinary Lagrange problem with fixed end points. Then the functions ψ_i of §2 are independent of x, and $p=n$. We shall understand that the indices used here have the following ranges: $i, j, k, l=1, \cdots, n$; $\alpha=1, \cdots, m$; $r=m+1, \cdots, n$. From the definitions of the functions ψ_i, S_{ij}, and G, we obtain the following relations:

$$\psi_{kz_i}\phi_{iy_j'} = \delta_{kj}, \tag{13}$$

$$\psi_{kz_j}\phi_{jy_i} = -\psi_{ky_i}, \tag{14}$$

$$\psi_{jy_l}(x) - \int_x^v S_{jk}(v, t)dt\,\psi_{ky_l}(x) = -S_{jl}(v, x), \tag{15}$$

$$G_{z_i} = (l_0 f_{y_j'} + c_j)\psi_{jz_i}, \tag{16}$$

$$G_{y_j} = l_0 f_{y_j} + (l_0 f_{y_k'} + c_k)\psi_{ky_j}. \tag{17}$$

The analogue of the Euler-Lagrange equations may be written

$$\begin{aligned} G_{z_i}(x) + \psi_{kz_i}(x)\int_x^{x_2}\left[G_{y_k}(t) - G_{y_j}(t)\int_x^t S_{jk}(t, v)dv\right]dt &= \lambda_i(x), \\ \lambda_r(x) &= 0. \end{aligned} \tag{18}$$

If we multiply equations (18) by $\phi_{iy_l'}$ and add, use equations (13) and (16), and interchange the order of integration in the double integral, we find

$$l_0 f_{y_l'} + c_l + \int_x^{x_2} G_{y_l}(t)dt - \int_x^{x_2}\int_t^{x_2} G_{y_j}(v)S_{jl}(v, t)dvdt = \lambda_\alpha \phi_{\alpha y_l'}. \tag{19}$$

Also if we multiply equations (18) by ϕ_{iy_l} and add, we find with the help of equations (14), (15) and (16),

$$\int_x^{x_2} G_{y_j}(v)S_{jl}(v, x)dv = (l_0 f_{y_j'} + c_j)\psi_{jy_l} + \lambda_\alpha \phi_{\alpha y_l}. \tag{20}$$

Combining equations (20) and (17) with (19) we find

$$l_0 f_{y_l'} + c_l + \int_x^{x_2}(l_0 f_{y_l} - \lambda_\alpha \phi_{\alpha y_l})dt = \lambda_\alpha \phi_{\alpha y_l'}$$

which may be written in the familiar form

$$F_{y_i'} = \int_{x_2}^{x} F_{y_i} dx - c_i \tag{21}$$

by setting $F = l_0 f - \lambda_\alpha \phi_\alpha$.

If we apply the inverse transformation in the more general problem considered in §1, we find in place of equations (21),

$$F_{y_i'}(x) = \int_{x_2}^{x} \left[F_{y_i}(s) + \int_{s}^{x_2} F_{u_\gamma}(t) P_{\gamma y_i}(t, s) dt \right.$$
$$\left. + F_{u_\gamma}(s) P_{\gamma y_i'}(s, x) \right] ds - c_i.$$

UNIVERSITY OF CHICAGO,
CHICAGO, ILL.

CONTRIBUTIONS TO THE THEORY OF TRANSFORMATIONS OF NETS IN A SPACE S_n*

BY

V. G. GROVE

1. Introduction

Let there be given a surface S in euclidean space of $n \geqq 3$ dimensions. Suppose that through each point x of S there passes a line g of a congruence G. The developables of G intersect S in a net of curves N. We have called such a net a *C net*.†

Let $\bar{S}$ be another surface in the same space S_n, in one-to-one point correspondence with S, corresponding points lying on the lines g of G. The developables of G intersect $\bar{S}$ in a C net of curves $\bar{N}$. The two nets N and $\bar{N}$ are said to be *in relation*‡ C.

The tangent planes to S and $\bar{S}$ intersect in a line h. If the points of h are each equidistant from the corresponding points of S and $\bar{S}$, we shall say that the nets N and $\bar{N}$ are *in relation E*.

We propose in this paper to develop a theory of the relations defined above which is independent of the dimension of the space S_n for $n \geqq 3$.

Let the coordinates of the point x on S be $x_1, x_2, \cdots, x_n$, the coordinates of the corresponding point $\bar{x}$ on $\bar{S}$ be $\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_n$, and the direction cosines of the line g joining them be $\lambda_1, \lambda_2, \cdots, \lambda_n$, where

$$\sum_{i=1}^{n} \lambda_i^2 = 1.$$

Let the parametric curves on S and $\bar{S}$ be chosen as the curves of the given nets N, $\bar{N}$ on these surfaces. The pairs of functions $(x, \bar{x})$ and the number pair $(1, 1)$ are solutions of a system of differential equations of the form§

$$\begin{aligned} \bar{x}_u &= m x_u - A x + A \bar{x}, \\ \bar{x}_v &= n x_v - B x + B \bar{x}. \end{aligned} \tag{1}$$

* Presented to the Society, April 14, 1933; received by the editors January 15, 1933.

† V. G. Grove, *The transformation C of nets in hyperspace*, these Transactions, vol. 33 (1931), pp. 733–741. Hereafter referred to as C.

‡ C, p. 733.

§ C, p. 734.

The coordinates of the point $\bar{x}$ are of the form

$$\bar{x} = x + \lambda\delta. \tag{2}$$

The pairs of functions (x, λ) are solutions of the following system of differential equations:

$$\begin{aligned} \lambda_u &= \mu x_u + \alpha\lambda, \\ \lambda_v &= \nu x_v + \beta\lambda, \end{aligned} \tag{3}$$

wherein

$$\mu = (m - 1)/\delta, \quad \alpha = -\mu E^{1/2} \cos \theta^{(u)}, \quad E = \sum_{i=1}^{n} x_{iu}^2,$$

$$\nu = (n - 1)/\delta, \quad \beta = -\nu G^{1/2} \cos \theta^{(v)}, \quad G = \sum_{i=1}^{n} x_{iv}^2,$$

and $\theta^{(u)}$ and $\theta^{(v)}$ are the angles between g and the tangents to v=const. and u=const. respectively.

From (2) we find that

$$\begin{aligned} \bar{E}^{1/2} \cos \bar{\theta}^{(u)} &= E^{1/2} \cos \theta^{(u)} + \delta_u, \\ \bar{G}^{1/2} \cos \bar{\theta}^{(v)} &= G^{1/2} \cos \theta^{(v)} + \delta_v, \end{aligned} \tag{4}$$

wherein $\bar{E}$, $\bar{G}$ etc. bear the same relation to $\bar{N}$ as the corresponding quantities bear to N.

The focal points ξ and η of the line g have the coordinates

$$\xi = x - \lambda/\mu, \quad \eta = x - \lambda/\nu, \quad \mu\nu \neq 0. \tag{5}$$

If $\mu\nu = 0$ one or both of the families of developables of G are cylinders.

The tangent planes to S and $\bar{S}$ at x and $\bar{x}$ intersect in the line h determined by the two points

$$r = x - m x_u/A, \quad s = x - n x_v/B, \quad AB \neq 0. \tag{6}$$

If $AB = 0$, one or both of the curves of the net N are parallel to the corresponding curves of $\bar{N}$. The points r and s are equidistant from x and $\bar{x}$ if

$$A\delta + 2mE^{1/2} \cos \theta^{(u)} = 0, \quad B\delta + 2nG^{1/2} \cos \theta^{(v)} = 0. \tag{7}$$

We may readily verify that equations (7) are necessary and sufficient conditions that the nets N and $\bar{N}$ be in relation E if not both A and B are zero. Equations (7) are of the form

$$\delta_u + P\delta = Q, \quad \delta_v + P'\delta = Q',$$

wherein P, Q, P', Q' are independent of δ.

If we differentiate the first of equations (1) with respect to v and the second with respect to u, we find that if $m-n\neq 0$,

$$x_{uv} = ax_u + bx_v - Mx + M\bar{x}, \tag{8}$$

wherein a, b, M are defined by

$$\begin{aligned} (m-n)a &= B(m-1) - m_v, \quad (m-n)M = B_u - A_v, \\ (n-m)b &= A(n-1) - n_u. \end{aligned} \tag{9}$$

If $m-n=0$, we find that

$$B(m-1) - m_v = 0, \quad A(n-1) - n_u = 0. \tag{10}$$

In case that N is not conjugate, and C is not radial, we find from (8) that

$$\begin{aligned} \bar{x} &= x + (x_{uv} - ax_u - bx_v)/M, \\ \lambda &= (x_{uv} - ax_u - bx_v)/(\delta M). \end{aligned} \tag{11}$$

If the net N is conjugate, or if C is radial, the congruence G is not determined by the net N alone.

2. Congruences semi-normal to a net

A congruence G will be said to be *semi-normal to the net corresponding to the developables of the congruence* if the lines g of G are perpendicular to the tangents of one (only) of the families of curves of the net. In particular suppose that the line g is perpendicular to the tangent at x to the curve $v=$ const. Suppose that the transformation C is a transformation E. It follows from (2) and (3) that

$$A = \delta_u = 0.$$

Hence *if a congruence is semi-normal to the net N in which the developables of the congruence intersect the sustaining surface S of N, the congruence is semi-normal to any E transform of N*. Moreover *the distance between corresponding points x and $\bar{x}$ on the curves of N and $\overline{N}$ to whose tangents the lines g are normal, is a constant, and the tangent lines to these curves are parallel.*

3. Two-parameter families of lines normal to a surface

Let Γ be a two-parameter family of lines, such that through each point x of S there passes one and only one line l of Γ. Suppose furthermore that this line l is perpendicular to the tangent plane to S at x for all points x on S. We shall say that Γ is *normal to S*. Let the direction cosines of l be $l_1, l_2, \cdots, l_n$. It follows therefore that

$$\sum l x_u = 0, \quad \sum l x_v = 0.$$

Consider a curve C on S with parametric equations

$$u = u(t), \quad v = v(t).$$

Any point y on the tangent t to C at x has coordinates defined by an expression of the form

$$y = x + p(x_u u' + x_v v'), \quad u' = \frac{du}{dt}\cdot$$

As x moves along C the point y describes a curve, the direction cosines of whose tangent are proportional to expressions of the form

$$p(x_{uu}u'^2 + 2x_{uv}u'v' + x_{vv}v'^2) + L(x_u, x_v),$$

wherein $L(x_u, x_v)$ is a homogeneous linear function of the indicated arguments. The line l is perpendicular to the tangent to the locus of the point y if and only if C is an integral curve of the differential equation

$$Ddu^2 + 2D'dudv + D''dv^2 = 0, \tag{12}$$

wherein

$$D = \sum l x_{uu}, \; D' = \sum l x_{uv}, \; D'' = \sum l x_{vv}. \tag{13}$$

We shall call the net defined by (12), in case a net is so defined, *the A net of* Γ. We readily verify that *the line l is normal to the osculating plane at x of any curve of the A net of* Γ. *The A net of* Γ *is indeterminate in case* Γ *is normal to every plane of the two-osculating space* $S_{(2,0)}$ *of S at x*. If the parametric net is a conjugate net it follows that $D'=0$ for the A net of every two-parameter family of lines Γ normal to S.

Suppose now that Γ is a congruence G. Let the parametric curves be the curves in which the developables of G intersect S. Equations (3) may be written

$$\lambda_u = \mu x_u, \quad \lambda_v = \nu x_v. \tag{14}$$

It follows therefore, that, if C is not radial, the functions x and λ each satisfy differential equations of the Laplace type. Moreover

$$F = \sum x_u x_v = 0,$$
$$\mathcal{J} = \sum \lambda_u \lambda_v = 0.$$

Hence *if a congruence is normal to a surface its developables intersect the surface in an orthogonal conjugate net. Moreover the net of curves of the spherical indicatrix of G corresponding to the developables of G is an orthogonal conjugate net.*

If the parametric curves are not the curves in which the developables of G intersect S, the curves in which these developables do intersect S are defined by the differential equation

$$(ED' - FD)du^2 + (ED'' - GD)dudv + (FD'' - GD')dv^2 = 0. \tag{15}$$

We remark at this point that *a given surface S cannot possess a normal congruence unless it sustains an orthogonal conjugate net. Moreover it cannot possess more than one such normal congruence unless the developables of such other congruence intersect the surface in the same net* (15). The tangents to the curves of the A nets of such other congruences belong to the same involution, namely that determined by the tangents to the minimal curves and the tangents to the curves of the A net of the given normal congruence.

4. The radial transformation E

Suppose that the transformation C is radial. It follows that

$$m - n = 0.$$

Suppose also that C is an E transformation. From (7) and (10), we find that

$$\delta^2 = k^2(m - 1)^2/m, \tag{16}$$

wherein k is an arbitrary constant different from zero. Conversely if $m-n=0$, and equation (16) is satisfied, so also are equations (7). If r and $\bar{r}$ denote the distances from the points x and $\bar{x}$ respectively to the focal point of g, we find readily that

$$r\bar{r} = \delta^2 m/(m - 1)^2 = k^2.$$

Hence *if two nets are radial transforms in relation E, they are transforms of one another by a transformation by reciprocal radii, and conversely.*

Equations (1) for a transformation by reciprocal radii assume the following simple form:

$$\begin{aligned} \bar{x}_u &= k^2\mu^2 x_u - (x - \bar{x})\frac{\partial}{\partial u}\log(k^2\mu^2 - 1), \\ \bar{x}_v &= k^2\mu^2 x_u - (x - \bar{x})\frac{\partial}{\partial v}\log(k^2\mu^2 - 1). \end{aligned} \tag{17}$$

Equations (17) may readily be integrated. The solution for a proper choice of the constants of integration may be written in the form

$$\bar{x} = k^2\mu^2 x. \tag{18}$$

The lines joining x to $\bar{x}$ evidently pass through the origin. Moreover from (5) the quantity $-1/\mu$ is the distance from the point x to the fixed focal point of g. Hence

$$\bar{x} = \frac{k^2 x}{\sum x^2}\,.$$

These of course are the familiar formulas of a transformation by reciprocal radii.

MICHIGAN STATE COLLEGE,
EAST LANSING, MICH.

ON THE EQUATION $P(A, X)=0$ IN MATRICES*

BY
WILLIAM E. ROTH

In the present discussion we shall consider the solution of the equation

$$(1) \qquad P(A, X) \equiv \sum_{k=0}^{p} F_k(A)X^{p-k} = 0,$$

where A is a known $n\times n$ matrix, $F_k(\lambda)$ $(k=0, 1, \cdots, p)$ are polynomials† in the scalar variable λ, and X is the unknown $n\times n$ matrix. The equation is a special case under that of an earlier paper by the author in which the coefficient matrices are not polynomials in a given matrix, but are known $m\times n$ matrices.‡ With the restrictions upon the coefficients which we now impose, it is possible to establish inequalities limiting the degree and the number of the elementary divisors of $X-\mu I$, where X is a solution of (1). These inequalities depend upon a knowledge of the elementary divisors of $P(A, \mu)$ and of $A-\lambda I$, where μ and λ are scalar variables. Certain theorems below, particularly Theorems III and IV with appropriate changes, are valid for the more general equation of the type studied by the author and others.§

Solutions of (1) are taken up under the following hypotheses: (a) that X be a unilateral solution on the right (or left) of the polynomials $F_k(\lambda)$ $(k=0, 1, \cdots, p-1)$; (b) that X be a bilateral solution; and (c) that X be commutative with A. By means of the idea of transversion of matrices as defined in §III, we show the fundamental relationship which exists between solutions on the right and those on the left of (1), and between these and the bilateral solutions if such exist.

---

* Presented to the Society, November 30, 1928, and September 9, 1931; received by the editors October 7, 1932.

† Considerations which follow do not require the functions $F_k(\lambda)$ $(k=0, 1, \cdots, p)$ to be polynomials. In fact any functions which, together with at most their first $n-1$ derivatives, may be expanded into series of non-negative powers of λ are permissible, provided that the characteristic values of A lie within or on the circles of convergence of each of the series representing $F_k(\lambda)$ $(k=0, 1, \cdots, p)$ and their first $n-1$ derivatives. For information on such functions of matrices the reader may consult Hensel, *Über Potenzenreihen von Matrizen*, Journal für die reine und angewandte Mathematik, vol. 155, pp. 107–110; Sheffer, *A note on matrix power series*, American Mathematical Monthly, vol. 36 (1930), pp. 228–231.

‡ Roth, *On the unilateral equation in matrices*, these Transactions, vol. 32 (1930), pp. 61–80. This paper cites several articles on algebraic equations in matrices.

§ Roth, loc. cit.

I. Preliminary notions and lemmas

Definition. *If* $A(\lambda)=(a_{ij}(\lambda))$ $(i=1, 2, \cdots, r; j=1, 2, \cdots, s)$, *where* $a_{ij}(\lambda)$ *are polynomials in* λ, *if*

$$a_{ij}(\lambda) \equiv r_{ij}(\lambda), \bmod (\lambda - a)^n \qquad (i = 1, 2, \cdots, r; j = 1, 2, \cdots, s),$$

and if $R(\lambda)=(r_{ij}(\lambda))$, *then*

$$A(\lambda) \equiv R(\lambda), \qquad \bmod (\lambda - a)^n.$$

Definition. *If* $A(\lambda)$ *is an* $m\times m$ λ*-matrix whose elements are polynomials in* λ, *if*

$$A(\lambda) \equiv R(\lambda), \qquad \bmod (\lambda - a)^n,$$

and if the ith elementary divisor of $R(\lambda)$ *corresponding to the linear factor* $\lambda-a$ *is* $(\lambda-a)^{\alpha^{(i)}}$ $(i=1, 2, \cdots, \rho)$, *where* ρ *is the rank of* $R(\lambda)$, *then* $(\lambda-a)^{\alpha^{(i)}}$ $(i=1, 2, \cdots, \rho)$ *is the ith elementary divisor of* $A(\lambda)$ *and* ρ *its rank with respect to the modulus* $(\lambda-a)^n$.

Definition. *If* $A(\lambda)$ *is an* $m\times m$ λ*-matrix whose elements are polynomials in* λ, *if the ith elementary divisor of* $A(\lambda)$ *corresponding to the linear factor* $\lambda-a$ *is* $(\lambda-a)^{\alpha^{(i)}}$ $(i=1, 2, \cdots, r)$ *where* r *is the rank of* $A(\lambda)$, *and if*

$$\alpha^{(i)} < n \qquad (i = 1, 2, \cdots, \sigma),$$
$$\alpha^{(i)} \geqq n \qquad (i = \sigma + 1, \sigma + 2, \cdots, r);$$

then σ *is the reduced rank and* $(\lambda-a)^{\alpha^{(i)}}$ $(i=1, 2, \cdots, \sigma)$ *is the ith elementary divisor of* $A(\lambda)$ *with respect to the modulus* $(\lambda-a)^n$.

Plainly if r is the rank $A(\lambda)$ and if ρ and σ are respectively the rank and the reduced rank of $A(\lambda)$ with respect to the modulus $(\lambda-a)^n$, then $\sigma\leqq\rho\leqq r\leqq m$; moreover it should be noted that in either case all minors of order $k\leqq\sigma$ are divisible by $\prod_{i=1}^{k} (\lambda-a)^{\alpha^{(i)}}$ and that this kth determinant divisor may consequently be congruent to zero modulo $(\lambda-a)^n$, while the elementary divisors with respect to this modulus are not. In speaking of the elementary divisors of $A(\lambda)$ with respect to the modulus $(\lambda-a)^n$, it is not necessary to designate the linear factor, for it is always that occurring in the modulus. As a matter of convenience we shall still call $(\lambda-a)^{\alpha^{(k)}}$ the kth elementary divisor of $A(\lambda)$ even when $\alpha^{(k)}=0$. Thus if $|A(\lambda)|$ is prime to $\lambda-a$, then each of the m elementary divisors of $A(\lambda)$ with respect to the modulus $(\lambda-a)^n$ is unity.

Lemma I. *If* $A(\lambda)$ *is an* $m\times m$ *matrix having elements* $a_{ij}(\lambda)$ $(i, j=1, 2, \cdots, m)$ *which are polynomials in* λ, *and if the reduced rank of* $A(\lambda)$ *with respect to the modulus* $(\lambda-a)^n$ *is* σ *and its ith elementary divisor is* $(\lambda-a)^{\alpha^{(i)}}$ $(i=1, 2, \cdots, \sigma)$, *then two* $m\times m$ *matrices,* $P(\lambda)$ *and* $Q(\lambda)$, *of degree* $n-1$ *in* λ,

exist such that $P(a)$ and $Q(a)$ are non-singular matrices whose elements do not depend upon n but do depend upon a, and that

$$P(\lambda)A(\lambda)Q(\lambda) \equiv S(\lambda), \qquad \text{mod } (\lambda - a)^n,$$

where

$$S(\lambda) \equiv (s_{ij}(\lambda))$$

and

$$s_{ij}(\lambda) \equiv 0, \qquad \text{mod } (\lambda - a)^n \qquad (i \neq j),$$

$$s_{ii}(\lambda) \equiv (\lambda - a)^{\alpha^{(i)}}, \text{mod } (\lambda - a)^n \qquad (i = 1, 2, \cdots, \sigma),$$

$$s_{ii}(\lambda) \equiv 0, \qquad \text{mod } (\lambda - a)^n \quad (i = \sigma + 1, \sigma + 2, \cdots, m),$$

and

$$\alpha^{(i)} \leqq \alpha^{(i+1)} \qquad (i = 1, 2, \cdots, \sigma - 1).$$

According to a well known theorem* two non-singular $m \times m$ matrices, $T(\lambda)$ and $U(\lambda)$, $|T(\lambda)|$ and $|U(\lambda)|$ independent of λ, exist such that

$$T(\lambda)A(\lambda)U(\lambda) = \prod_{i=1}^{t} A_i(\lambda), \tag{2}$$

where $\lambda - a_i$ $(i=1, 2, \cdots, t)$ are the distinct linear factors in λ common to all r-rowed minors of $A(\lambda)$, where r is the rank of $A(\lambda)$, and where $A_h(\lambda) = (a_{ij}^{(h)}(\lambda))$ $(h=1, 2, \cdots, t)$ are such that

$$a_{ij}^{(h)}(\lambda) = 0 \qquad (i \neq j),$$

$$a_{ii}^{(h)}(\lambda) = (\lambda - a_h)^{\alpha_h^{(i)}} \qquad (i \leqq r),$$

$$a_{ii}^{(h)}(\lambda) = 0 \qquad (i > r).$$

Hence the kth composite elementary divisor of $A(\lambda)$, as usually defined, is

$$\prod_{h=1}^{t}(\lambda - a_h)^{\alpha_h^{(k)}} \qquad (k = 1, 2, \cdots, r).$$

Since $A_i(\lambda)$ $(i=1, 2, \cdots, t)$ are diagonal matrices, they are commutative one with another and (2) can be written in the form

$$T(\lambda)A(\lambda)U(\lambda) = A_k(\lambda)\left[\prod_{i=1}^{k-1} A_i(\lambda)\prod_{i=k+1}^{t} A_i(\lambda)\right]. \tag{3}$$

Now if the last $n-r$ zero elements in the principal diagonal of

$$A_i(\lambda) \qquad (i = 1, 2, \cdots, k - 1, k + 1, k + 2, \cdots, t)$$

be replaced by unit elements, the resulting matrix, $A_i'(\lambda)$ $(i=1, 2, \cdots,$

* Bôcher, *Introduction to Higher Algebra*, 1922, p. 91, and Theorem I, p. 94.

$k-1, k+1, k+2, \cdots, t$) may replace the corresponding matrix $A_i(\lambda)$ in the right member of (3), and this substitution will not affect the validity of this equation in that the product by $A_k(\lambda)$ leaves this member unchanged. The jth element $j \leqq r$ of the diagonal matrix

$$\prod_{i=1}^{k-1} A_i'(\lambda) \prod_{i=k+1}^{t} A_i'(\lambda) \text{ is } \prod_{i=1}^{k-1} (\lambda - a_i)^{\alpha_i^{(j)}} \prod_{i=k+1}^{t} (\lambda - a_i)^{\alpha_i^{(j)}}$$

and is a polynomial of degree $(\sum_{i=1}^{t} \alpha_i^{(j)}) - \alpha_k^{(j)}$ in $\lambda - a_k$ whose constant term is not zero. Hence the polynomial $v_{jk}(\lambda)$ exists such that

$$\prod_{i=1}^{k-1} (\lambda - a_i)^{\alpha_i^{(j)}} \prod_{i=k+1}^{t} (\lambda - a_i)^{\alpha_i^{(j)}} v_{jk}(\lambda) = 1 + (\lambda - a_k)^{n_k} w_{jk}(\lambda), \; j \leqq r,$$

where $w_{jk}(\lambda)$ is a polynomial in λ. The remaining $n-r$ elements of $\prod_{i=1}^{k-1} A_i'(\lambda) \prod_{i=k+1}^{t} A_i'(\lambda)$ are unit elements, hence we may let $v_{jk}(\lambda) \equiv 1$ for $r < j \leqq n$. Hence the diagonal matrix $V_k(\lambda)$, having as its jth element $v_{jk}(\lambda)$ as defined above, exists such that

$$\left[\prod_{i=1}^{k-1} A_i'(\lambda) \prod_{i=k+1}^{t} A_i'(\lambda)\right] V_k(\lambda) = I + (\lambda - a_k)^{n_k} W_k(\lambda),$$

where $|V_k(a_k)| \neq 0$ and $V_k(a_k)$ is again independent of n_k. Hence (3) becomes

$$T(\lambda) A(\lambda) U(\lambda) V_k(\lambda) = A_k(\lambda) [I + (\lambda - a_k)^{n_k} W_k(\lambda)],$$

where $U(a_k) V_k(a_k)$ is a non-singular matrix independent of n_k. Now if we let $a_k = a$ and let

$$T(\lambda) \equiv P(\lambda), \qquad \text{mod } (\lambda - a)^n,$$
$$U(\lambda) V_k(\lambda) \equiv Q(\lambda), \qquad \text{mod } (\lambda - a)^n,$$
$$A_k(\lambda) \equiv S(\lambda), \qquad \text{mod } (\lambda - a)^n,$$

we have demonstrated that $P(\lambda)$ and $Q(\lambda)$ satisfying the lemma exist.

LEMMA II. *If* $\Delta(\lambda) = (\delta_{ij}(\lambda))$ $(i, j = 1, 2, \cdots, n)$ *where*

$$\delta_{i,i+k}(\lambda) \equiv (\lambda - a)^{\alpha_k} d_k(\lambda) \qquad (k = 0, 1, \cdots, n-1),$$
$$\delta_{i,i-h}(\lambda) \equiv 0 \qquad (h = 1, 2, \cdots, n-1),$$

where $d_k(\lambda)$ $(k = 0, 1, \cdots, n-1)$ *are polynomials in* λ *and where* $d_0(a) \neq 0$ *and* $d_1(a) \neq 0$, *then the degree of the nth elementary divisor of* $\Delta(\lambda)$ *corresponding to the linear factor* $\lambda - a$

(1) *does not exceed* $n\alpha_0 - n + 1$ *if* $\alpha_0 > 1$ *and* $\alpha_1 > 0$,

(2) *does not exceed* $(n+1)/2$ *if* $\alpha_0 = 1$ *and* $\alpha_1 > 0$,

(3) *is equal to* $n\alpha_0$ *if* $\alpha_1 = 0$.

The proof of this lemma consists in seeking a lower bound to the degree of $\lambda - a$ as a divisor of all $(n-1)$st-order minors of $\Delta(\lambda)$. The determinant

$|\Delta(\lambda)|$ has the divisor $(\lambda-a)^{n\alpha_0}$, and none of higher degree in $\lambda-a$ if the lemma is satisfied; hence the difference between the lower bound so sought and $n\alpha_0$ is an upper bound for the degree of the nth elementary divisor of $\Delta(\lambda)$.

The minor $\delta_{ii}(\lambda)$ has the factor $(\lambda-a)^{(n-1)\alpha_0}$, in $\lambda-a$; that of $\delta_{i,i+k}(\lambda)$ $(k=1, 2, \cdots, n-1)$ is identically zero, whereas the minor of $\delta_{i,i-h}(\lambda)$ $(h=1, 2, \cdots, n-1)$ is $[(\lambda-a)^{\alpha_0}d_0(\lambda)]^{n-h-1} D_h(\lambda)$, where $D_h(\lambda)$ is the minor of order h obtained by dropping the first column and last $n-h-1$ columns and the last $n-h$ rows of $\Delta(\lambda)$.

Now

$$D_h(\lambda) = (\lambda - a)^{\alpha_1}d_1(\lambda)D_{h-1}(\lambda) - (\lambda - a)^{\alpha_0+\alpha_2}d_0(\lambda)d_2(\lambda) + \cdots \tag{4}$$
$$\pm (\lambda - a)^{(h-1)\alpha_0+\alpha_h}d_0^{\,h-1}(\lambda)d_h(\lambda) \qquad (h = 1, 2, \cdots, n - 1),$$

and $D_0(\lambda)=1$. If $\alpha_0>1$ and $\alpha_1>0$, we can show by mathematical induction on the basis of the recurrence relation (4), that $D_h(\lambda)$ has at least the factor $(\lambda-a)^h$, hence the minor of any element $\delta_{i,i-h}(\lambda)$ $(h=1, 2, \cdots, n-1)$ has at least the divisor $(\lambda-a)^{(n-h-1)\alpha_0+h}$ and that of lowest degree among them occurs for $h=n-1$; hence all $(n-1)$st-order minors of $\Delta(\lambda)$ have at least the factor $(\lambda-a)^{n-1}$ in common and for this case the degree of the nth elementary divisor of $\Delta(\lambda)$ corresponding to the linear factor $\lambda-a$ is at most $n\alpha_0-n+1$.

If $\alpha_0=1$ and $\alpha_1>0$, $D_1(\lambda)$ and $D_2(\lambda)$ have at least the factor $\lambda-a$. Then from (4) it readily follows that $(\lambda-a)^{h/2}$ or $(\lambda-a)^{(h+1)/2}$ are factors of $D_h(\lambda)$ according as h is an even or an odd integer. Hence we can infer that the minor of $\delta_{i,i-h}(\lambda)$ $(h=1, 2, \cdots, n-1)$ is divisible by $(\lambda-a)^{n-h-1+h/2}$ or by $(\lambda-a)^{n-h-1+(h+1)/2}$ according as h is even or odd. The divisor of lowest degree occurs for $h=n-1$, and is $(\lambda-a)^{n/2}$ or $(\lambda-a)^{(n-1)/2}$ according as n is an even integer or an odd integer. That is, if $\alpha_0=1$ and $\alpha_1>0$ the degree of the nth elementary divisor of $\Delta(\lambda)$ does not exceed $n/2$ or $(n+1)/2$ according as n is an even or an odd integer.

The third part of the lemma is evident, for the minor of $\delta_{n,1}(\lambda)$ is prime to $\lambda-a$, since all terms of its expansion save $d_1^{\,n-1}(\lambda)$ have this factor if $\alpha_1=0$. Hence in this case the nth elementary divisor of $\Delta(\lambda)$ corresponding to $\lambda-a$ is $(\lambda-a)^{n\alpha_0}$.

II. The unilateral solution

Let the normal form of A be given by $\overline{A}=(A_{ij})$, where

$$A_{ij} = 0 \qquad (i \neq j),$$
$$A_{ii} = A_i \qquad (i = 1, 2, \cdots, r),$$

and A_i $(i=1, 2, \cdots, r)$ is an $m_i\times m_i$ matrix, $\sum_{i=1}^{r} m_i=n$, the elements of whose principal diagonal are a_i and those in the diagonal directly above are

m_i-1 unit elements and the remaining $(m_i-1)^2$ elements of A_i are zeros. Hence $A-\lambda I$ has the simple elementary divisors $(\lambda-a_i)^{m_i}$ $(i=1, 2, \cdots, r)$ and the non-singular matrix Q exists such that

$$A = Q\overline{A}Q^{-1}. \tag{5}$$

Moreover let (1) have the solution X on the right whose normal form is $\overline{X}=(X_{ij})$ $(i, j=1, 2, \cdots, s)$, where

$$\begin{aligned} X_{ij} &= 0 && (i \neq j), \\ X_{jj} &= X_j = x_jI_j + D_j && (j = 1, 2, \cdots, s), \end{aligned}$$

where I_j is the $n_j \times n_j$ unit matrix, x_j is a scalar constant and D_j is the $n_j\times n_j$ matrix, whose elements are all zeros save those in the kth row and $(k+1)$st column $(k=1, 2, \cdots, n_j-1)$ which are unities. Thus we may write $D_j^0=I_j$ and $D_j^k=0$, $k\geqq n_j$. The matrix $X-\mu I$ has the elementary divisors $(\mu-x_j)^{n_j}$ $(j=1, 2, \cdots, s)$ and the non-singular matrix R exists such that

$$X = R\overline{X}R^{-1}. \tag{6}$$

On substituting for A and X in (1) by means of (5) and (6), and noting that Q and R are non-singular matrices, we obtain

$$\textstyle\sum_{k=0}^{p} F_k(\overline{A})T\overline{X}^{p-k} = 0, \tag{7}$$

where $T=Q^{-1}R$. In this equation $\overline{X}$ and R (hence T) are the unknowns. In fact $T\overline{X}T^{-1}$ is a solution of $P(\overline{A}, X)=0$; on the other hand if X is a solution of $P(\overline{A}, X)=0$, then $Q^{-1}XQ$ is a solution of (1).

Let $T=(T_{ij})$, where T_{ij} $(i=1, 2, \cdots, r; j=1, 2, \cdots, s)$ is an $m_i\times n_j$ matrix; then from (7) we readily obtain the rs equations

$$\textstyle\sum_{k=0}^{p} F_k(A_i)T_{ij}X_j^{p-k} = 0 \quad (i = 1, 2, \cdots, r; j = 1, 2, \cdots, s), \tag{8}$$

which must be satisfied by the rs independent matrices T_{ij}. Each of these equations provides a means of computing the corresponding T_{ij}, and consequently T, provided the matrices $X_j (j=1, 2, \cdots, s)$ were known. We shall seek restrictions upon x_j of X_j and upon its order n_j.

Now from $X_j=x_jI_j+D_j$ we have

$$X_j^h = \textstyle\sum_{k=0}^{h} \binom{h}{k} x_j^{h-k} D_j^k,$$

and consequently (8), for A_i and X_j, becomes

$$\textstyle\sum_{k=0}^{h} \dfrac{P_{0k}(A_i, x_j)}{k!} T_{ij}D_j^k = 0, \tag{9}$$

where

$$P_{hk}(\lambda, \mu) = \frac{\partial^{h+k}}{\partial\lambda^h\partial\mu^k}P(\lambda, \mu).$$

This equation must be satisfied by the sub-matrix T_{ij} of T, in order that (1) have a solution whose characteristic matrix, $X-\mu I$, has the elementary divisor $(\mu-x_j)^{n_j}$, where $(\lambda-a_i)^{m_i}$ is an elementary divisor of $A-\lambda I$.

Indicate the $m_i\times 1$ matrix formed by the $(k+1)$st column of T_{ij} by the *script* letter $\mathfrak{T}_{ij}^{(k)}$; then

$$T_{ij} = (\mathfrak{T}_{ij}^{(0)}, \mathfrak{T}_{ij}^{(1)}, \cdots, \mathfrak{T}_{ij}^{(n_j-1)})$$

and

$$T_{ij}D_j^k = (0, \cdots, 0, \mathfrak{T}_{ij}^{(0)}, \mathfrak{T}_{ij}^{(1)}, \cdots, \mathfrak{T}_{ij}^{(n_j-k-1)});$$

that is, the multiplication of T_{ij} on the right by D_j^k moves the first n_j-k columns of T_{ij} k spaces to the right and replaces the evacuated spaces by k zero columns. Hence from (9) we readily obtain the equations

$$\sum_{h=0}^{k} \frac{P_{0h}(A_i, x_j)}{h!}\mathfrak{T}_{ij}^{(k-h)} = 0 \qquad (k = 0, 1, \cdots, n_j - 1).$$

Multiply these for $k=0, 1, \cdots, n_j-1$ respectively by $1, \mu-x_j, \cdots, (\mu-x_j)^{n_j-1}$ and add the results; the single equation

$$P(A_i, \mu)\mathfrak{T}_{ij}(\mu) \equiv 0, \qquad \operatorname{mod} (\mu - x_j)^{n_j}, \tag{10}$$

is thus obtained, where

$$\mathfrak{T}_{ij}(\mu) = \sum_{h=0}^{n_j-1}(\mu - x_j)^h\mathfrak{T}_{ij}^{(h)} = T_{ij}\begin{pmatrix} 1 \\ \mu - x_j \\ \vdots \\ (\mu - x_j)^{n_j-1}\end{pmatrix} \tag{11}$$

is consequently an $m_i\times 1$ matrix whose elements are polynomials of degree n_j-1 in $\mu-x_j$. We shall henceforth concentrate upon equation (10) instead of (8).

From (10) and (11), we see that $\mathfrak{T}_{ij}(x_j)=\mathfrak{T}_{ij}^{(0)}=0$, if $|P(A_i, x_j)|\neq 0$; hence also $\mathfrak{T}_{ij}^{(1)}=0$ and so on, under the same hypothesis. That is, $T_{ij}=0$, if $|P(A_i, x_j)|\neq 0$. Now not all T_{ij} $(i=1, 2, \cdots, r)$ can be zero, nor can all T_{ij} $(j=1, 2, \cdots, s)$ be zero else T would have n_j zero columns or m_i zero rows and in either case would be a singular matrix. Hence $|P(A_i, x_j)|=0$ for at least one pair of values of i and j; the necessary and sufficient condition that such be the case is that $P(a_i, x_j)=0$. We have proved the theorem.

THEOREM I. *If the characteristic matrix $A-\lambda I$, of A, have the elementary divisors $(\lambda-a_i)^{m_i}$ $(i=1, 2, \cdots, r)$, where $\sum_{i=1}^{r} m_i=n$, and if $P(A, X)=0$ have a solution, X, on the right (or left) whose characteristic matrix, $X-\mu I$, has the elementary divisors $(\mu-x_j)^{n_j}$, where $\sum_{j=1}^{s} n_j=n$, then every equation $P(a_i, \mu)=0$ $(i=1, 2, \cdots, r)$ must be satisfied by at least one of the numbers x_j $(j=1, 2, \cdots, s)$ and every equation $P(\lambda, x_j)=0$ $(j=1, 2, \cdots, s)$ must be satisfied by at least one of the numbers a_i $(i=1, 2, \cdots, r)$.*†

The above theorem shows where and how the characteristic values x_j of a solution of (1) must be sought and consequently gives us some knowledge of the sub-matrices X_j $(j=1, 2, \cdots, s)$. For more definite information regarding them, we shall seek restrictions upon n_j, in addition to that we already know, namely that $\sum_{j=1}^{s} n_j=n$ in order that the non-singular matrix $T=(T_{ij})$ may exist. Such is given by the following theorems.

THEOREM II. *If X is a solution of the polynomial equation $P(A, X)=0$, and if $X-\mu I$ has the elementary divisors $(\mu-x)^{\nu_1}, (\mu-x)^{\nu_2}, \cdots, (\mu-x)^{\nu_k}$ corresponding to the linear factor $\mu-x$, then $(\mu-x)^{\nu_1+\nu_2+\cdots+\nu_k}$ is a factor of $|P(A, \mu)|$; moreover if $X-\mu I$ has the elementary divisors $(\mu-x_j)^{n_j}$ $(j=1, 2, \cdots, s)$, then $\prod_{j=1}^{s} \{P(\lambda, x_j)\}^{n_j}$ must be an exact multiple of $|A-\lambda I|$.*

It is known that if $P(A, X)=0$, then $|P(A, \mu)|$ is divisible by $|X-\mu I|$‡, and this determinant in turn is the product of all its elementary divisors; hence the first part of the theorem is proved. Similarly if X is a solution of (1), then A satisfies the same equation, where we regard X as the known matrix, and consequently $|P(\lambda, X)|$ must be divisible by $|A-\lambda I|$. But we can readily show that

$$|P(\lambda, X)| = \prod_{j=1}^{s} |P(\lambda, X_j)| = \prod_{j=1}^{s} \{P(\lambda, x_j)\}^{n_j}.$$

Hence the second part of the theorem is proved.

The restrictions placed upon n_j by this theorem are not very severe; nevertheless the first part of the theorem places an upper bound upon n_j and the second part places a lower bound upon n_j $(j=1, 2, \cdots, s)$. The following results are far more restrictive and quite as easily applied in particular examples as are the above.

THEOREM III. *If $P(A, \mu)$ is of rank ρ_j with respect to the modulus $(\mu-x_j)^{n_j}$ and if the ρ_jth elementary divisor of $P(A, \mu)$ with respect to the same modulus*

---

† This theorem is in part a special case of one proved elsewhere, Roth, loc. cit., Theorem I, p. 65.

‡ Roth, loc. cit., Corollary I, p. 66.

is $(\mu-x_j)^{\alpha_j^{(\rho_j)}}$, *then the equation* (1) *may have the solution X whose characteristic matrix*, $X-\mu I$, *has the elementary divisor* $(\mu-x_j)^{n_j}$ *only if*

$$\alpha_j^{(\rho_j)} \geqq n_j + \rho_j - n;$$

and if $\alpha_j^{(\rho_j)} < n_j$, *that is, if the reduced rank and the rank of* $P(A,\mu)$ *with respect to the modulus* $(\mu-x_j)^{n_j}$ *are the same, then* $X-\mu I$ *has the elementary divisor* $(\mu-x_j)^{n_j}$ *at most k times only if*

$$\alpha_j^{(\rho_j)} \geqq n_j - \frac{n-\rho_j}{k}.$$

If

$$\frac{P_{h,0}(a_i,\mu)}{h!} \equiv 0, \bmod (\mu - x_j)^{n_j} \qquad (h = 0, 1, \cdots, \sigma_{ij} - 1),$$

$$\frac{P_{h,0}(a_i,\mu)}{h!} \equiv p_{ij}^{(h)}(\mu), \bmod (\mu - x_j)^{n_j} \quad (h = \sigma_{ij}, \sigma_{ij} + 1, \cdots, m_i - 1);$$

then $P(A_i, \mu)$ is of rank $\rho_{ij}=m_i-\sigma_{ij}$ with respect to the modulus $(\mu-x_j)^{n_j}$ and equation (10) reduces to the following non-homogeneous system of ρ_{ij} equations in the ρ_{ij} unknowns $t_{ij}^{(\sigma+h)}(\mu)$ $(h=0, 1, \cdots, \rho_{ij}-1)$†:

$$\begin{pmatrix} p^{(\sigma)}(\mu) & p^{(\sigma+1)}(\mu) & \cdots & p^{(m_i-1)}(\mu) \\ 0 & p^{(\sigma)}(\mu) & \cdots & p^{(m_i-2)}(\mu) \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \cdots & p^{(\sigma)}(\mu) \end{pmatrix} \begin{pmatrix} t^{(\sigma)}(\mu) \\ t^{(\sigma+1)}(\mu) \\ \vdots \\ t^{(m_i-1)}(\mu) \end{pmatrix} \tag{12}$$
$$= (\mu - x_j)^{n_j} \begin{pmatrix} q^{(\sigma)}(\mu) \\ q^{(\sigma+1)}(\mu) \\ \vdots \\ q^{(m_i-1)}(\mu) \end{pmatrix},$$

where $q^{(k)}(\mu)$ $(k=\sigma, \sigma+1, \cdots, m_i-1)$ are arbitrary polynomials in μ. The equation (10) imposes no restrictions upon $t^{(0)}(\mu), t^{(1)}(\mu), \cdots, t^{(\sigma-1)}(\mu)$, hence each of the first σ_{ij} rows of T_{ij} has n_j arbitrary elements and its remaining rows must be such that (11) and (12) are satisfied. From (12) we see at once that

$$t^{(\sigma+h)}(\mu) = \frac{(\mu - x_j)^{n_j} M_h(\mu)}{[p^{(\sigma)}(\mu)]^{m_i-\sigma-h}} \quad (h = 0, 1, \cdots, \rho_{ij} - 1),$$

† In this equation and in the remainder of the proof of this lemma we suppress the subscripts i and j of $p_{ij}^{(k)}(\mu)$ of σ_{ij} and of $t_{ij}^{(\sigma+h)}(\mu)$ save where ambiguity may arise.

where $M_h(\mu)$ is a linear combination of minors of order $\rho_{ij}-h-1$ of $P(A_i, \mu)$ and consequently has the $(\rho_{ij}-h-1)$th determinant divisor in $\mu-x_j$ of this matrix as its divisor. We shall now seek a lower bound to the degree of $\mu-x_j$ as a divisor of $t^{(k)}(\mu)$ $(k=\sigma, \sigma+1, \cdots, m_i-1)$.

Let $p^{(\sigma)}(\mu)$ have the factor $(\mu-x_j)^{\epsilon_{ij}}$ and $t^{(k)}(\mu)$ $(k=\sigma, \sigma+1, \cdots, m_i-1)$ have the factor $(\mu-x_j)^{\tau_{ij}^{(k)}}$ and let neither have a divisor of higher degree than these in $\mu-x_j$. Moreover, let the kth elementary divisor of $P(A_i, \mu)$ be $(\mu-x_j)^{\alpha_{ij}^{(k)}}$ $(k=1, 2, \cdots, \rho_{ij})$. Then the determinant divisor of all minors of order g of $P(A_i, \mu)$ is $\prod_{k=1}^{g} (\mu-x_j)^{\alpha_{ij}^{(k)}}$, $g \leqq \rho_{ij}$, and

$$(m_i - \sigma_{ij})\epsilon_{ij} = \rho_{ij}\epsilon_{ij} = \sum_{k=1}^{\rho_{ij}} \alpha_{ij}^{(k)}.$$

$M_h(\mu)$ has the factor $\mu-x_j$ at least $\sum_{k=1}^{\rho_{ij}-h-1} \alpha_{ij}^{(k)}$ times, and $\tau_{ij}^{(\sigma+h)}$ must satisfy the inequality

$$\tau_{ij}^{(\sigma+h)} \geqq \eta_j - \sum_{k=1}^{\rho_{ij}} \alpha_{ij}{}^{k} + h\epsilon_{ij} + \sum_{k=1}^{\rho_{ij}-h-1} \alpha_{ij}{}^{k},$$

or

$$\tau_{ij}^{(\sigma+h)} \geqq n_j + h\epsilon_{ij} - \sum_{k=\rho_{ij}-h}^{\rho_{ij}} \alpha_{ij}^{(k)} \qquad (h = 0, 1, \cdots, \rho_{ij} - 1). \tag{13}$$

This inequality evidently establishes a lower bound for the number of zero elements in the $(\sigma+h)$th row of T_{ij}, for if $\tau_{ij}^{(\sigma+h)}=k$ then the elements in the first k columns and the $(\sigma+h)$th row of T_{ij} are zero. The least value the right member of the inequality may have for any h occurs for $h=0$, that is, $\tau_{ij}^{(\sigma+h)} \geqq n_j-\alpha_{ij}^{(\rho_{ij})}$ $(h=0, 1, \cdots, \rho_{ij}-1)$. Therefore T_{ij} has only zero elements in at least the first $n_j-\alpha_{ij}^{(\rho_{ij})}$ columns of the last ρ_{ij} rows, whereas the first $m_i-\rho_{ij}=\sigma_{ij}$ rows have arbitrary elements as was pointed out above. Since

$$P(\overline{A}, \mu) = \begin{pmatrix} P(A_1, \mu) & 0 & \cdots & 0 \\ 0 & P(A_2, \mu) & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & P(A_r, \mu) \end{pmatrix},$$

the rank of $P(\overline{A}, \mu)$, hence of $P(A, \mu)$, with respect to the modulus $(\mu-x_j)^{n_j}$ is ρ_j, where $\rho_j=\sum_{i=1}^{r} \rho_{ij}$ and ρ_{ij} is the rank of $P(A_i, \mu)$ with respect to the same modulus. Moreover, if the ρ_j numbers $\alpha_{ij}^{(k)}$ $(k=1, 2, \cdots, \rho_{ij}; i=1, 2, \cdots, r)$ be rearranged in an ascending sequence

$$\alpha_j^{(1)} \leqq \alpha_j^{(2)} \leqq \cdots \leqq \alpha_j^{(\rho_j)},$$

then $(\mu-x_j)^{\alpha_j^{(k)}}$ is the kth elementary divisor of $P(\overline{A}, \mu)$ and of $P(A, \mu)$ with

respect to the modulus $(\mu-x_j)^{n_j}$ and $\alpha_j^{\rho_j}$ is the greatest of the numbers $\alpha_{ij}^{(\rho_{ij})}$ $(i=1, 2, \cdots, r)$.

According to (10)

$$P(A, \mu)\mathfrak{T}_j(\mu) \equiv 0, \bmod (\mu - x_j)^{n_j}, \tag{14}$$

where

$$\mathfrak{T}_j(\mu) = \begin{Vmatrix} \mathfrak{T}_{1j}(\mu) \\ \mathfrak{T}_{2j}(\mu) \\ \cdot \\ \cdot \\ \mathfrak{T}_{rj}(\mu) \end{Vmatrix} = \begin{Vmatrix} T_{1j} \\ T_{2j} \\ \cdot \\ \cdot \\ T_{rj} \end{Vmatrix} \begin{Vmatrix} 1 \\ \mu - x_j \\ \cdot \\ \cdot \\ (\mu - x_j)^{n_j-1} \end{Vmatrix} = T_j \begin{Vmatrix} 1 \\ \mu - x_j \\ \cdot \\ \cdot \\ (\mu - x_j)^{n_j-1} \end{Vmatrix}.$$

The matrix T_j has $\sum_{i=1}^{r}{}' (m_i-\rho_{ij}) = n-\rho_j$ rows of arbitrary elements and the remaining rows have only zero elements in the first $n_j-\alpha_j^{(\rho_j)}$ columns. Hence $\alpha_j^{(\rho_j)}$ must equal or exceed $n_j-n+\rho_j$ else T_j is of rank less than n_j and T would be singular. This proves the first part of the theorem. Now if $n_j>\alpha_j^{(\rho_j)}$ the reduced rank of $P(A, \mu)$ is ρ_j and if $X-\mu I$ have the elementary divisor $(\mu-x_j)^{n_j}$ repeated k times then T has k matrices T_j all having the same $n-\rho_j$ rows of arbitrary elements. Each T_j has at least $n_j-\alpha_j^{(\rho_j)}$ zero columns in the remaining rows. Hence $k(n_j-\alpha_j^{(\rho_j)})$ cannot exceed $n-\rho_j$ else the corresponding kn_j columns of T are of rank less than kn_j and T would be singular. This proves the final part of the theorem.

The second part of the theorem may be stated as follows:

Corollary I. *If the rank ρ_j of $P(A, \mu)$ with respect to the modulus $(\mu-x_j)^{n_j}$ is equal to the reduced rank with respect to the same modulus and if $(\mu-x_j)^{\alpha_j^{(\rho_j)}}$ is its ρ_jth elementary divisor, then the characteristic matrix, $X-\mu I$, of a solution of* (1) *cannot have the elementary divisor $(\mu-x_j)^{n_j}$ more than $(n-\rho_j)/(n_j-\alpha_j^{(\rho_j)})$ times.*

Plainly if n_j be taken sufficiently large the rank of $P(A, \mu)$ with respect to the modulus $(\mu-x_j)^{n_j}$ is equal to the rank, r, of $P(A, \mu)$ in the usual sense; that is, all minors of order $r+1$ and above are identically zero whereas those of order r are not all identically zero and in this case by Theorem III we have $n_j \leqq \alpha_j^{(r)}-n+r$, where $(\mu-x_j)^{\alpha_j^{(r)}}$ is the rth elementary divisor of $P(A, \mu)$ corresponding to the linear factor $\mu-x_j$. Hence:

Corollary II. *If $P(A, \mu)$ is of rank r and if the rth elementary divisor of $P(A, \mu)$ corresponding to the linear factor $\mu-x_j$ is $(\mu-x_j)^{\alpha_j^{(r)}}$, then no solution X of $P(A, X)=0$ exists whose characteristic matrix, $X-\mu I$, has an elementary divisor corresponding to the linear factor $\mu-x_j$ whose degree exceeds*

$$\alpha_j^{(r)} - n + r.$$

The following corollary is at once evident from the foregoing.

COROLLARY III. *If $P(A, \mu)$ is of rank $r<n$, then $X-\mu I$, where X is a solution of* (1), *may have the elementary divisor $(\mu-x)^k$, where x is an arbitrary parameter, only if $k\leqq n-r$.*

In this case, where x is arbitrary, the rth elementary divisor of $P(A, \mu)$ corresponding to $\mu-x$ is unity, and $\alpha^{(r)}=0$. A more complete discussion of this case is given in the paper cited above†, where the method of computing the matrix corresponding to T is covered in some detail.

THEOREM IV. *If $P(A, \mu)$ has the reduced rank ρ with respect to the modulus $(\mu-x)^\nu$, and if $P(A, X)=0$, then the number of elementary divisors of $X-\mu I$ corresponding to the same linear factor $\mu-x$ and whose degree equals or exceeds ν cannot exceed $n-\rho$.*

If $P(A, \mu)$ has the elementary divisors $(\mu-x)^{\alpha^{(k)}}$ $(k=1, 2, \cdots, \rho)$, where $\alpha^{(1)}\leqq\alpha^{(2)}\leqq \cdots \leqq\alpha^{(\rho)}<\nu$, and if the remaining elementary divisors of $P(A, \mu)$ corresponding to the same linear factor $\mu-x$ are all of degree equal to or greater than ν, then according to Lemma I there exist matrices $R(\mu)$ and $S(\mu)$ such that $|R(x)|\neq 0$ and $|S(x)|\neq 0$ and that

$$R(\mu)P(A, \mu)S(\mu) \equiv Q(\mu), \qquad \text{mod } (\mu - x)^\nu,$$

where $Q(\mu)=(q_{ij}(\mu))$ $(i, j=1, 2, \cdots, n)$ is given by

$$\begin{aligned} q_{ij}(\mu) &\equiv 0, & &\text{mod } (\mu - x)^\nu & &(i \neq j),\\ q_{ii}(\mu) &\equiv (\mu - x)^{\alpha^{(i)}}, & &\text{mod } (\mu - x)^\nu & &(i \leqq \rho),\\ q_{ii}(\mu) &\equiv 0, & &\text{mod } (\mu - x)^\nu & &(i > \rho).\end{aligned}$$

By (5) $P(A, \mu)=QP(\overline{A}, \mu)Q^{-1}$, hence

$$R(\mu)QP(\overline{A}, \mu)Q^{-1}S(\mu) = Q(\mu), \text{ mod } (\mu - x)^\nu,$$

and (14) becomes

$$Q(\mu)S^{-1}(\mu)Q\mathfrak{T}'(\mu) \equiv 0, \text{ mod } (\mu - x)^\nu,$$

where T' is an $n\times\nu$ matrix formed of ν adjacent columns of T and $\mathfrak{T}'(\mu)$ is the $n\times 1$ matrix, whose elements are polynomials of degree $\nu-1$ in $\mu-x$, and $\mathfrak{T}'(x)$ is the first of these ν columns of T'. From this equation we see that the element of the kth row of the $n\times 1$ matrix $S^{-1}(\mu)\, Q\mathfrak{T}'(\mu)$ is divisible by $(\mu-x)^{\nu-\alpha^{(k)}}$, if $k\leqq\rho$, and is prime to $\mu-x$, if $k>\rho$. Consequently the $n\times 1$ matrix

$$S^{-1}(x)Q\mathfrak{T}'(x) = U$$

† Roth, loc. cit., §3.

has ρ zero elements in the first ρ rows and arbitrary elements in the remaining $n-\rho$ rows. Now if $X-\mu I$ has the k elementary divisors $(\mu-x)^{\nu_i}$, $\nu_i \geqq \nu$ $(i=1, 2, \cdots, k)$, and if X is a solution of (1), then for each $(\mu-x)^{\nu_i}$ we must have

$$S^{-1}(x)Q\mathfrak{T}_i'(x) = U_i,$$

where U_i has zero elements in at least the first ρ rows. The reduced rank of $P(A, \mu)$ with respect to the modulus $(\mu-x)^{\nu_i}$, $\nu_i \geqq \nu$, cannot be less than ρ, and $S^{-1}(x)$ is not dependent upon the degree of the modulus $(\mu-x)^{\nu}$. Consequently

$$S^{-1}(x)Q(\mathfrak{T}_1'(x), \mathfrak{T}_2'(x), \cdots, \mathfrak{T}_k'(x)) = (U_1, U_2, \cdots, U_k).$$

The rank of $S^{-1}(x)Q$ is n and the rank of the right member is at most $n-\rho$; hence in order that the k columns $\mathfrak{T}_i'(x)$ $(i=1, 2, \cdots, k)$ of T may form a matrix of rank k, k cannot exceed $n-\rho$. The theorem here demonstrated is more general than Corollary I, but if $n_j-\alpha_j^{(\rho_j)} \geqq 2$, the latter offers the more restrictive bound upon the number of equal elementary divisors that $X-\mu I$ may have.

THEOREM V. *If $P(a_i, \mu)=0$ has the root x_j of multiplicity β_{ij}, and if $P(\lambda, x_j)=0$ has the root a_i of multiplicity γ_{ij}, if $A-\lambda I$ has the elementary divisors $(\lambda-a_i)^{m_i}$ $(i=1, 2, \cdots, r)$ and if $X-\mu I$ has the elementary divisors $(\mu-x_j)^{n_j}$ $(j=1, 2, \cdots, s)$ where X is a solution of $P(A, X)=0$, then at least one $n_j (j=1, 2, \cdots, s)$ must equal or exceed the corresponding n_{ij} for each value of i $(i=1, 2, \cdots, r)$, where*

$$n_{ij} = \frac{m_i - 1}{\gamma_{ij} - 1} \qquad (\beta_{ij} > 1, \gamma_{ij} > 1),$$

$$n_{ij} = 2m_i - 1 \qquad (\beta_{ij} > 1, \gamma_{ij} = 1),$$

$$n_{ij} = \frac{m_i}{\gamma_{ij}} \qquad (\beta_{ij} = 1, \gamma_{ij} \geqq 1),$$

$$n_{ij} = \infty \qquad (\beta_{ij} = \gamma_{ij} = 0).$$

Under the hypotheses of this theorem neither $P(a_i, \mu)$ $(i=1, 2, \cdots, r)$ nor $P(\lambda, x_j)$ $(j=1, 2, \cdots, s)$ is identically zero. Hence the rank of $P(\lambda, X)$, where X is a solution of $P(A, X)=0$, and of $P(A, \mu)$ is n. Now if we regard A as a solution of (1), where X is the known matrix, then according to Corollary II, m_i is less than or equal to $\beta_i^{(n)}$, where $(\lambda-a_i)^{\beta_i^{(n)}}$ is the nth elementary divisor of $P(\lambda, X)$. Now at least one of the matrices $P(\lambda, X_j)$ $(j=1, 2, \cdots, s)$ must have $(\lambda-\alpha_i)^{\beta_i^{(n)}}$ as its n_jth elementary divisor corresponding to the linear factor $\lambda-a_i$. Lemma II gives us a means of computing an upper bound

to the degree of the n_jth elementary divisor of $P(\lambda, X_j)$ $(j=1, 2, \cdots, s)$. For if we set $\Delta(\lambda)=P(\lambda, X_j)$, then

$$\delta_{i,i+k}(\lambda) = \frac{P_{0,k}(\lambda, x_j)}{k!} \qquad (k = 0, 1, \cdots, n_j - 1)$$

and if $(\mu-x_j)^{\gamma_{ij}}$, $\gamma_{ij}>1$, is a factor of $P(a_i, \mu)$, then $\delta_{i,i+1}(\lambda)=P_{0,1}(\lambda, x_j)$ will have $\lambda-a_i$ as a factor; on the other hand if $\gamma_{ij}=1$, $P_{0,1}(\lambda, x_j)$ is prime to $\lambda-a_i$. Let $P(\lambda, x_j)$ have the factor $(\lambda-a_i)_{\beta^{ij}}$ but not one of higher degree in $\lambda-a_i$; then according to Lemma II and Theorem III, m_i cannot exceed every m_{ij}, where

$$m_{ij} = \beta_{ij}n_j - n_j + 1 \qquad (\beta_{ij} > 1, \gamma_{ij} > 1),$$

$$m_{ij} = \frac{n_j + 1}{2} \qquad (\beta_{ij} = 1, \gamma_{ij} > 1),$$

$$m_{ij} = \beta_{ij}n_j \qquad (\beta_{ij} \geqq 1, \gamma_{ij} = 1),$$

$$m_{ij} = 0 \qquad (\beta_{ij} = 0, \gamma_{ij} = 0).$$

Hence not all n_j $(j=1, 2, \cdots, s)$ can be less than the numbers n_{ij} defined in the theorem. When $\beta_{ij}=\gamma_{ij}=0$, then $|P(A_i, x_j)|\neq 0$ and the corresponding $T_{ij}=0$, and we must take $m_{ij}=0$ since not all T_{ij} $(j=1, 2, \cdots, s)$ can be zero. Similarly n_{ij} must be taken sufficiently large in case $\beta_{ij}=\gamma_{ij}=0$.

According to the theorem above if $\beta_{ij}=\gamma_{ij}=1$, and if $\beta_{hj}=0$, $h\neq i$, and $\gamma_{ik}=0$, $k\neq j$, we must have $n_j\geqq m_i$, for $n_{ij}=m_i$ and $n_{ik}=\infty$, $k\neq j$. The m_ith elementary divisor of $P(A_i, \mu)$ corresponding to the linear factor $\mu-x_j$ is $(\mu-x_j)^{m_i}$ because of Lemma I, and $P(A_h, \mu)$, $h\neq i$, has only the elementary divisors unity corresponding to the same linear factor if $\beta_{hj}=0$. Hence the elementary divisor of highest degree of $P(A, \mu)$ corresponding to $\mu-x_j$ is $(\mu-x_j)^{m_i}$. That is, by Theorem III, $n_j\leqq m_i$. Consequently under the hypotheses here laid down $n_j=m_i$; and the equation $P(A_i, \mu)\ \mathfrak{T}_{ij}(\mu)\equiv 0$, mod $(\mu-x_j)^{m_i}$, has a solution such that $|T_{ij}|\neq 0$. We have consequently proved the following corollary.

Corollary IV. *If $A-\lambda I$ has the elementary divisors $(\lambda-a_i)^{m_i}$ $(i=1, 2, \cdots, r)$ such that $a_k\neq a_i$, $k\neq i$, and if the equations $P(a_i, \mu)=0$ $(i=1, 2, \cdots, r)$ have the distinct simple roots x_{ij} $(j=1, 2, \cdots, p_i)$ such that a_i is a simple root of each of the equations $P(\lambda, x_{ij})=0$ $(j=1, 2, \cdots, p_i)$ and that $P(a_k, x_{ij})\neq 0$, $k\neq i$ $(j=1, 2, \cdots, p_i)$, then $P(A, X)=0$ has $\prod_{i=1}^{r} p_i$ solutions, X, such that $X-\mu I$ has the elementary divisors $(\mu-x_{ij})^{m_i}$ $(i=1, 2, \cdots, r)$.*

The numbers x_{ij} in the elementary divisor $(\mu-x_{ij})^{m_i}$ $(i=1, 2, \cdots, r)$ can be chosen in p_i ways and all are distinct, for if $x_{ij}=x_{hk}$, $i\neq h$, then $P(a_h, x_{ij}) = P(a_h, x_{hk})=0$, which is contrary to hypothesis.

III. The bilateral solution

Definition. *If* $B=(b_{ij})$ $(i=1, 2, \cdots, \alpha; j=1, 2, \cdots, \beta)$, *then* $B' = (b_{\beta-j-1,\alpha-i+1})$ *is the transverse of* B.

The transverse of a matrix is obtained by reflecting its elements with respect to a line at right angles to that with respect to which the transpose of the matrix is obtained. For example, if

$$B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{pmatrix}, \quad B' = \begin{pmatrix} b_{32} & b_{22} & b_{12} \\ b_{31} & b_{21} & b_{11} \end{pmatrix}.$$

The following theorems hold:

The transverse of the sum of two or more matrices is equal to the sum of their transverses.

The transverse of the product of two or more matrices is equal to the product of their transverses taken in reverse order; $(AB)' = B'A'$.

If $A = aI + D$, *where* $D = (\delta_{ij})$ *and*

$$\delta_{i,i+1} = 1 \qquad (i = 1, 2, \cdots, n-1),$$
$$\delta_{ij} = 0 \qquad (i + 1 \neq j),$$

then $A' = A$.

Definition. *If* $B = (B_{ij})$, *where* B_{ij} $(i=1, 2, \cdots, r; j=1, 2, \cdots, s)$ *are* $\alpha_i \times \beta_j$ *matrices such that* $\sum_{i=1}^{r}\alpha_i = \alpha$, $\sum_{j=1}^{s}\beta_j = \beta$, *then the* $\beta \times \alpha$ *matrix*

$$B^* = (B_{ji}'),$$

where B_{ij}' *is the transverse of* B_{ij}, *is the compound transverse of* B *with respect to the sub-matrices* B_{ij}.

The compound transverse of a matrix depends upon the way it is divided into sub-matrices. The following theorems hold.

The compound transverse of the sum of two or more matrices is equal to the sum of the transverses of the addend matrices, provided all addend matrices are divided into sub-matrices in the same way.

If $B = (B_{ij})$ *and* $C = (C_{jk})$, *where* B_{ij} $(i=1, 2, \cdots, r; j=1, 2, \cdots, s)$ *are* $\alpha_i \times \beta_j$ *matrices and* C_{jk} $(j=1, 2, \cdots, s; k=1, 2, \cdots, t)$ *are* $\beta_j \times \gamma_k$ *matrices, such that* $\sum_{i=1}^{r}\alpha_i = \alpha$, $\sum_{j=1}^{s}\beta_j = \beta$, *and* $\sum_{k=1}^{t}\gamma_k = \gamma$, *then the compound transverse of* AB *is the* $\gamma \times \alpha$ *matrix obtained by multiplying the compound transverse of* B *on the right by the compound transverse of* A; *that is,*

$$(AB)^* = B^*A^*.$$

If $\overline{A}$ and $\overline{X}$ are the matrices in the normal forms as given in §II, *and if transversion of them is made with respect to their sub-matrices A_i $(i=1, 2, \cdots, r)$ and $X_j (j=1, 2, \cdots, s)$ respectively, then $\overline{A}^*=\overline{A}$ and $\overline{X}^*=\overline{X}$.*

If A is an $n\times n$ matrix, then the elementary divisors of $A-\lambda I$ are identical with those of $(A-\lambda I)^*$ for any division of $A-\lambda I$ into sub-matrices, and identical with those of $A^*-\lambda I$ provided the transversion of A is made with respect to its sub-matrices A_{ij} $(i, j=1, 2, \cdots, r)$ such that A_{ii} are all square matrices of order n_i and $\sum_{i=1}^{r} n_i=n$.

If B is a non-singular $n\times n$ matrix and if B^{-1} is its inverse, then for every division of B into sub-matrices there exists a corresponding division of B^{-1} into sub-matrices such that for B and B^{-1} so divided

$$(B^*)^{-1} = (B^{-1})^*.$$

If $B=(B_{ij})$ and $B^{-1}=(C_{ji})$ where B_{ij} are $\alpha_i\times\beta_j$ matrices and C_{ji} are $\beta_j\times\alpha_i$ matrices, the theorem is satisfied provided $\sum_{i=1}^{r}\alpha_i=\sum_{j=1}^{s}\beta_j=n$.

The idea of transversion and compound transversion of matrices as defined above enables us to determine the relationship of a solution on the right of (1) to one having the same normal form on the left, and their relation to the bilateral solution of the same equation.

THEOREM VI. *If $A=Q\overline{A}Q^{-1}$ and $X=R\overline{X}R^{-1}$, where $\overline{A}$ and $\overline{X}$ are the normal forms of A and X as defined in* §II, *if $A-\lambda I$ and $X-\mu I$ have the elementary divisors $(\lambda-a_i)^{m_i}$ $(i=1, 2, \cdots, r)$ and $(\mu-x_j)^{n_j}$ $(j=1, 2, \cdots, s)$ respectively, where $\sum_{i=1}^{r} m_i=\sum_{j=1}^{s} n_j=n$, and if X is a solution on the right of*

$$(1) \qquad P(A, X) = \sum_{k=0}^{p} F_k(A)X^{p-k} = 0,$$

then

$$X_1 = R_1\overline{X}R_1^{-1}$$

is a solution of

$$(15) \qquad \sum_{k=0}^{p} X^{p-k}F_k(A) = 0,$$

provided

$$R_1 = QQ^*(R^*)^{-1},$$

where Q and R are divided into sub-matrices of order $m_i\times n$ and $n_j\times n$ respectively.

If we assume that (15) has a solution on the left whose characteristic matrix $X_1-\mu I$ has the elementary divisors $(\mu-x_j)^{n_j}$ $(j=1, 2, \cdots, s)$ then

R_1 must exist such that $X_1 = R_1\overline{X}R_1^{-1}$. Hence by a procedure parallel to that of §II, we obtain from (15) the following equation:

$$\sum_{k=0}^{p} \overline{X}^{p-k} U F_k(\overline{A}) = 0, \tag{16}$$

where $U = R_1^{-1}Q$.

Now if we take the compound transverses of the members of (7) with respect to the sub-matrices A_i $(i = 1, 2, \cdots, r)$, X_j $(j = 1, 2, \cdots, s)$ and T_{ij} of $\overline{A}$, $\overline{X}$, and T respectively we have

$$\sum_{k=0}^{p} \overline{X}^{p-k} T^* F_k(\overline{A}) = 0.$$

That is $U = T^*$ satisfies (16) provided T satisfies (7), and similarly any U satisfying (16) is such that U^* satisfies (7). Hence

$$U = R_1^{-1}Q = T^* = R^*(Q^*)^{-1},$$

or $R_1 = QQ^*(R^*)^{-1}$ according to the theorems on the transversion of matrices. This proves the theorem.

In order that X be a bilateral solution of (1) it suffices but is not necessary that $R_1 = QQ^*(R^*)^{-1} = R$; in other words, that $RR^* = QQ^*$. The following theorem holds.

THEOREM VII. *In order that X be a bilateral solution of* (1) *it is necessary and sufficient that this equation have a solution $X = R\overline{X}R^{-1}$ on the right such that* (7) *is satisfied by T_1 and $T_2 = (T_{ij}^{(2)})$, not necessarily distinct, and such that $T_1T_2^* = I$, where T_2^* is the compound transverse of T_2 with respect to the sub-matrices $T_{ij}^{(2)}$ of order $m_i \times n_j$.*

If $X = R\overline{X}R^{-1}$ is a bilateral solution of (1), then $T_1 = Q^{-1}R$ satisfying (7) exists and $U = R^{-1}Q$ satisfying (16) exists and U^* also satisfies (7). That is, T_2, some solution of (7), is U^* or $(R^{-1}Q)^*$. It is sometimes possible that $T_1T_1^*$ cannot be a unit matrix, but if in T_2 we permit the parametric elements of T_1 to take another set of values, then it is possible for T_1 to be the inverse of T_2^* where T_2 is so taken.

IV. SOLUTIONS COMMUTATIVE WITH A

Little that is general can be said regarding the solutions of (1) which are commutative with A besides that already demonstrated to hold for the unilateral and bilateral solutions of the same equation. But with certain restrictions upon either A or X or on both we can derive such results on commutative matrices as are set forth in the following theorems.

THEOREM VIII. *If $AX=XA$, if $A-\lambda I$ has the elementary divisors $(\lambda-a_i)^{m_i}$ $(i=1, 2, \cdots, r)$, where $a_i\neq a_j$, $i\neq j$, and*

$$m_1 \geqq m_2 \geqq \cdots \geqq m_r,$$

and if $X-\mu I$ has the elementary divisors $(p-x_j)^{n_j}$ $(j=1, 2, \cdots, s)$, where

$$n_1 \geqq n_2 \geqq \cdots \geqq n_s,$$

but where $x_j (j=1, 2, \cdots, s)$ are not necessarily distinct; then

$$\sum_{k=1}^{h} n_k \leqq \sum_{k=1}^{h} m_k \qquad (h = 1, 2, \cdots, r)$$

and $s\geqq r$.

We shall here use the notation of §II with the understanding that the submatrices A_i $(i=1, 2, \cdots, r)$ and $X_j (j=1, 2, \cdots, s)$ of $\overline{A}$ and $\overline{X}$ are so ordered that their orders m_i and n_j respectively form non-increasing sequences of numbers. This is in no sense a restriction upon A nor upon X.

Since $AX=XA$, we have from (5) and (6) that

$$\overline{A}T\overline{X}T^{-1} = T\overline{X}T^{-1}\overline{A},$$

where $T=R^{-1}Q$. Therefore $T\overline{X}T^{-1}$ is commutative with $\overline{A}$, but the most general matrix commutative with $\overline{A}$, where all a_i $(i=1, 2, \cdots, r)$ are distinct, is $K=(K_{hk}\rfloor$, where

$$K_{hk} = 0 \qquad (h \neq k),$$
$$K_{hh} = c_0 I_h + c_1 D_h + \cdots + c_{m_h-1} D_h^{m_h-1} \quad (h = 1, 2, \cdots, r),$$

and where D_h is an $m_h \times m_h$ matrix having only unit elements in the diagonal immediately above the principal diagonal and having zero elements in the remaining $m_h^2 - m_h + 1$ places.† Hence

$$T_{ij}X_j = K_{ii}T_{ij} \qquad (i = 1, 2, \cdots, r; j = 1, 2, \cdots, s),$$

and

$$T_{ij}x_j + T_{ij}D_j = c_0 T_{ij} + \left\{c_1 D_i + c_2 D_i^2 + \cdots + c_{m_i-1} D_i^{m_i-1}\right\} T_{ij}.$$

If $c_0\neq x_j$, then $T_{ij}=0$, but not all T_{ij} $(i=1, 2, \cdots, r)$ may be zero else T would have n_j zero columns and would be singular. Hence we can assume that $c_0=x_j$ and the equation above reduces to

$$T_{ij}D_j = (c_1 D_i + c_2 D_i^2 + \cdots + c_{m_i-1} D_i^{m_i-1}) T_{ij}.$$

The multiplication of T_{ij} by D_j on the right moves the columns of T_{ij} one

† Kreis, *Contributions à la Théorie des Systèmes Linéaires*, Zurich Thesis, 1906. Hilton, *Linear Substitutions*, 1914, pp. 112–118.

space to the right, and the multiplication of T_{ij} on the left by D_i^k moves the rows of T_{ij} up k spaces. Because of this fact, whether c_1 is zero or not, T_{ij} has at least n_j-h zero elements in the first n_j-h places of the (m_i-h+1)st row $(h=1, 2, \cdots, m_i)$. If $n_j>m_i$, then at least the first n_j-m_i columns of T_{ij} have only zero elements. If any n_j exceeds every m_i $(i=1, 2, \cdots, s)$, then T will have at least one column of zero elements and this is impossible, consequently the largest n_j cannot exceed the largest m_i or

$$n_1 \leqq m_1.$$

This completes the first step of a mathematical induction proof. Now suppose we have shown that

$$\sum_{i=1}^{h} n_i \leqq \sum_{i=1}^{h} m_i, \tag{17}$$

and let

$$T = \begin{pmatrix} T_h & T_h^{(2)} \\ T_h^{(1)} & T_h^{(3)} \end{pmatrix},$$

where

$$\begin{aligned}
T_h &= (T_{ij}) & &(i, j = 1, 2, \cdots, h),\\
T_h^{(1)} &= (T_{ij}) & &(i = h+1, h+2, \cdots, r; j = 1, 2, \cdots, h),\\
T^{(2)} &= (T_{ij}) & &(i = 1, 2, \cdots, h; j = h+1, h+2, \cdots, s),\\
T^{(3)} &= (T_{ij}) & &(i = h+1, h+2, \cdots, r; j = h+1, h+2, \cdots, s).
\end{aligned}$$

If $n_{h+1} \leqq m_{h+1}$, then from (17) we have at once

$$\sum_{i=1}^{h+1} n_i \leqq \sum_{i=1}^{h+1} m_i$$

and the theorem holds; however if $n_{h+1}>m_{h+1}$, further proof is required. The number of zero columns in $T_{h+1}^{(1)}$ is at least equal to $\sum_{i=1}^{h+1}(n_i-m_{h+1})$, for if $n_{h+1}>m_{h+1}$, then every $n_i (i=1, 2, \cdots, h+1)$ exceeds m_j for $j \geqq h+1$ and all T_{ij} in $T_{h+1}^{(1)}$ have at least the first m_i-m_{h+1} columns of zero elements. The number of non-zero rows in T_{h+1} in the same columns where $T_{h+1}^{(1)}$ has only zero elements is equal to at most $\sum_{i=1}^{h+1}(m_i-m_{h+1})$. The rank of the first $\sum_{i=1}^{h+1} n_i$ columns of T will be less than $\sum_{i=1}^{h+1} n_i$ unless

$$\sum_{i=1}^{h+1}(n_i - m_{h+1}) \leqq \sum_{i=1}^{h+1}(m_i - m_{h+1}),$$

and the theorem is proved by mathematical induction for all h less than or equal to r or s. Since

$$\sum_{i=1}^{r} m_i = \sum_{i=1}^{s} n_j = n,$$

and if the inequality (17) holds, then it is not possible for r to be less than s. The theorem is proved.

COROLLARY V. *If $AX = XA$, if $A - \lambda I$ has the elementary divisors $(\lambda - a_i)^{m_i}$ $(i = 1, 2, \cdots, r)$, where $a_i \neq a_k$, $i \neq k$, and*

$$m_1 \geqq m_2 \geqq \cdots \geqq m_r,$$

and if $X - \mu I$ has the elementary divisors $(\mu - x_j)^{n_j}$ $(j = 1, 2, \cdots, s)$, where $x_j \neq x_h$, $j \neq h$, and

$$n_1 \geqq n_2 \geqq \cdots \geqq n_s,$$

then

$$m_i = n_i \qquad (i = 1, 2, \cdots, r),$$

and $r = s$.

This corollary is a direct consequence of the theorem above, for from it we find that

$$\sum_{i=1}^{h} n_i \leqq \sum_{i=1}^{h} m_i \text{ and } r \geqq s \qquad (h = 1, 2, \cdots, s);$$

and because $x_j \neq x_k$, $j \neq k$, that

$$\sum_{i=1}^{h} m_i \leqq \sum_{i=1}^{h} n_i \text{ and } r \leqq s \qquad (h = 1, 2, \cdots, r).$$

In order that these inequalities hold simultaneously, n_i must equal m_i and r must equal s.

The above theorem and corollary have obvious application to the solution of the equation $P(A, X) = 0$ for X commutative with A. However, matrices A and X such that $A - \lambda I$ and $X - \mu I$ have the elementary divisors $(\lambda - a_i)^{m_i}$ and $(\mu - x_i)^{m}$ respectively are not necessarily commutative, so that even when Corollary IV of the preceding section is satisfied it is not a simple matter to show that the solutions, whose existence is there established, are also commutative with A.

EXTENSION DIVISION,
UNIVERSITY OF WISCONSIN,
MILWAUKEE, WIS.

A SPECIAL INTEGRAL FUNCTION*

BY

R. E. A. C. PALEY

1. Some years ago Collingwood and Valiron† proposed the problem of whether there could exist an integral function whose minimum modulus on every circle $|z|=r$ is bounded, but possessing no asymptotic paths. By an asymptotic path we mean a continuous path tending to infinity along which the value of the function tends to a limit.

In this paper I show how to construct such a function. It is obtained by considering the well known Weierstrassian non-differentiable function

$$\sum_{n=0}^{\infty} c^n z^{a^n}$$

where $c(1<c<2)$ and a (an integer) are suitably chosen. We may observe that, if a is large enough, the Weierstrassian function possesses no asymptotic paths which tend to the boundary $|z|=1$, while, for sufficiently small c, its minimum modulus on circles $|z|=r<1$ is bounded, and every point of the unit circle is an essential singularity for the function.

2. Consider the function

$$F_N(z) = \left(\sum_{n=0}^{N} c^n z^{a^n} \right) \exp \left\{ - \left(\frac{z}{1-a^{-N}} \right)^{2^N a^N} \right\}$$

where $c>1$ (say $c=3/2$), and a is large. We first show that the minimum modulus of $F_N(z)$ on any circle $|z|=r$ is bounded, independently of N. Clearly, for $r>1$, $F_N(r)$ does not exceed

$$(1) \qquad Bc^N r^{a^N} \exp\left(-r - e^{B2^N}\right),$$

where B, here and in the sequel, denotes an absolute positive constant (it may denote a different constant in different contexts). For $r \geqq 1$, $N \geqq 1$, the expression (1) does not exceed a fixed constant, and thus it is sufficient to consider $F_N(z)$ with $|z| \leqq 1$.

Let r be a fixed number less than 1. We consider the value of $F_N(re^{i\theta})$ where θ is chosen according to the following rules. We first stipulate that

* Presented to the Society, June 23, 1933; received by the editors February 4, 1933. This paper was written by the late R. E. A. C. Paley in its present form; proof was read by J. D. Tamarlsin.

† E. F. Collingwood and G. Valiron, *Theorems concerning an analytic function which is bounded upon a simple curve passing through an isolated essential singularity*, Proceedings of the London Mathematical Society, (2), vol. 26 (1927), pp. 169–184; p. 182.

$a^N\theta \equiv 0 \pmod{2\pi}$, so that also $2^N a^N \theta \equiv 0 \pmod{2\pi}$, and thus the second factor

$$\exp\left\{-\left(\frac{z}{1-a^{-N}}\right)^{2^N a^N}\right\}$$

of $F_N(z)$ is real and less than 1 in modulus. There are now a possible reduced values of $a^{N-1}\theta \pmod{2\pi}$. We choose that one which makes

$$\left|\sum_{n=N-1}^{N} c^n r^{a^n} \exp(ia^n\theta)\right|$$

a minimum. We now choose that one of the reduced values of $A^{N-2}\theta$ (mod 2π) which makes

$$\left|\sum_{n=N-2}^{N} c^n r^{a^n} \exp(ia^n\theta)\right|$$

a minimum, and so on. We show that if a is sufficiently large the resulting value of

$$\left|\sum_{n=0}^{N} c^n r^{a^n} \exp(ia^n\theta)\right|$$

will not exceed a fixed constant independent of N. The argument is almost identical with that given in an earlier paper.* We have at the first stage a possible values of

$$\left|\sum_{n=N-1}^{N} c^n r^{a^n} \exp(ia^n\theta)\right|. \tag{2}$$

There is one of these for which the angle between the lines joining the point $c^N r^{a^N}$ to the origin and to the point

$$\sum_{n=N-1}^{N} c^n r^{a^n} \exp(ia^n\theta)$$

is less than or equal to π/a. Then the value of (2) can be seen by elementary geometry to lie between

$$c^{N-1} r^{a^{N-1}} \text{ and } c^N r^{a^N} \sec(\pi/a) - c^{N-1} r^{a^{N-1}},$$

which, if a is sufficiently large ($c = 3/2$), is certainly not greater than $c^{N-1} r^{a^{N-1}}$. Thus

$$\min_{a^N\theta\equiv 0} \left|\sum_{n=N-1}^{N} c^n r^{a^n} \exp(ia^n\theta)\right| \leq c^{N-1} r^{a^{N-1}}.$$

* R. E. A. C. Paley, *On some problems connected with Weierstrass's non-differentiable function*, Proceedings of the London Mathematical Society, (2), vol. 31 (1930), pp. 301–328; Theorem I, pp. 304–308.

Having now fixed the reduced value of $a^{N-1}\theta$ (mod 2π), we look for the minimum value of

$$(3) \qquad \left| \sum_{n=N-2}^{N} c^n r^{a^n} \exp\,(ia^n\theta) \right|.$$

There is certainly one value for the expression (3), such that the angle between the lines joining the point

$$\left| \sum_{n=N-1}^{N} c^n r^{a^n} \exp\,(ia^n\theta) \right|$$

to the origin and to the point

$$\sum_{n=N-2}^{N} c^n r^{a^n} \exp\,(ia^n\theta)$$

is less than or equal to π/a. Thus the value of (3) lies between

$$c^{N-2} r^{a^{N-2}} \text{ and } \left| \sum_{n=N-1}^{N} c^n r^{a^n} \exp\,(ia^n\theta) \right| \sec\left(\frac{\pi}{a}\right) - c^{N-2} r^{a^{N-2}},$$

and, if a is sufficiently large, it does not exceed $c^{N-2}r^{a^{N-2}}$. An inductive process will now show that

$$\min_{a^N\theta\equiv 0} \left| \sum_{n=0}^{N} c^n r^{a^n} \exp\,(ia^n\theta) \right| \leqq r \leqq 1,$$

and we have shown that, for all values of r, the minimum modulus of $F_N(z)$ on $|z| = r$ is less than an absolute constant.

The derivative $F_N'(z)$ of $F_N(z)$ is

$$(4) \quad \exp\left\{ -\left(\frac{z}{1-a^{-N}}\right)^{2^N a^N} \right\} \left\{ \sum_{n=0}^{N} (ac)^n z^{a^n} - \frac{2^N a^N}{z} \left(\frac{z}{1-a^{-N}}\right)^{2^N a^N} \left(\sum_{n=0}^{N} c^n z^{a^n} \right) \right\}.$$

Now suppose that a is sufficiently large, and that $1-3a^{-N} \leqq r \leqq 1-2a^{-N}$. Then the expression (4) is majorized* by the single term

$$(5) \qquad a^N c^N z^{a^N - 1}.$$

Indeed the single term (5) exceeds in modulus

$$a^N c^N (1 - 3a^{-N})^{a^N - 1} \geqq \frac{1}{32} a^N c^N, \quad a \geqq 6,$$

* See, e.g., G. H. Hardy, *Weierstrass' non-differentiable function*, these Transactions, vol. 17 (1916), pp. 301–332.

while the difference between the terms (4) and (5) is not greater in modulus than

$$\exp\left(\frac{1-2a^{-N}}{1-a^{-N}}\right)^{2^N a^N}\left\{\sum_{n=0}^{N-1}(ac)^n + 2\cdot 2^N a^N\left(\frac{1-2^{-N}}{1-a^{-N}}\right)^{2^N a^N}\left(\sum_{n=0}^{N}c^n\right)\right\}$$

$$+ a^N c^N\left\{\exp\left(\frac{1-2a^{-N}}{1-a^{-N}}\right)^{2^N a^N} - 1\right\} \leqq 10^{-6}a^N c^N$$

if a and N are sufficiently large.

3. We now write

$$F(z) = \sum_{k=1}^{\infty} f_k(u_k), \quad u_k = \left(\frac{z}{R_k}\right)^{\alpha_k},$$

and set, for abbreviation,

$$f_k(u) \equiv F_{N_k}(u), \; \alpha_k = a^{\lambda_k}, \; \beta_k = a^{N_k+\lambda_k},$$

where $\lambda_1=0$, $\lambda_{k+1}=2(N_k+\lambda_k)$, while N_k, R_k $(k=1, 2, \cdots)$ remain to be chosen. We write $N_1=R_1=1$ and give an inductive method for choosing N_k, R_k for $k>1$. Suppose that we have already chosen $N_1, \cdots, N_{k-1}, R_1, \cdots, R_{k-1}$. Since first $F_N(z)=O(|z|)$ for small z uniformly in N, we may choose R_k so large that

$$(6) \qquad |f_k(u_k)| \leqq 2^{-n}, |z| \leqq R_{k-n}, \; n = 1, 2, \cdots, k-1,$$

whatever the value of N_k may be. Next since, for $|z| \leqq R/2$ we have, uniformly in N_k and R,

$$\frac{d}{dz} f_k\left[\left(\frac{z}{R}\right)^{\alpha_k}\right] = O(|z|^{\alpha_k - 1}R^{-\alpha_k}),$$

we may also assume that R_k is so large that, whatever the value of N_k may be, we have

$$(7) \qquad \left|\frac{d}{dz} f_k(u_k)\right| \leqq 2^{-n}10^{-6}, \; |z| \leqq R_{n-k}, \quad n = 1, 2, \cdots, k-1.$$

This finally fixes R_k. We now choose $N_k > k$ so great that

$$(8) \qquad \beta_k c^{N_k} R_k^{-1} \geqq 10^6 \max_{|z| \leqq R_k} \left|\frac{d}{dz} \sum_{m=1}^{k-1} f_m(u_m)\right|.$$

We next observe that, for $-(2a)^{-N_k}\pi/4 \leqq \theta \leqq (2a)^{-N_k}\pi/4$, $r \geqq 1$,

$$| f_k(z) | \leqq \left| Bc^{N_k} r^{a^{N_k}} \exp\left(\frac{r}{1 - a^{-N_k}}\right)^{2^{N_k} a^{N_k}} e^{\pi i/4}\right\}\right|$$

$$= Bc^{N_k} r^{a^{N_k}} \exp\left\{- \; 2^{-1/2}\left(\frac{r}{1 - a^{-N_k}}\right)^{2^{N_k} a^{N_k}}\right\} \leqq Bc^{N_k} \exp\left\{- (Be)^{2^{N_k}}\right\}.$$

For $-a^{-\lambda_{k+1}}\pi \leqq \theta \leqq a^{-\lambda_{k+1}}\pi$, the argument of u_k is $\alpha_k\theta$, and thus in modulus does not exceed

$$a^{\lambda_k - \lambda_{k+1}}\,\pi = a^{-2N_k - \lambda_k}\pi \leqq (2a)^{-N_k}\pi/4;$$

whence, on the range $|z| \geqq R_k$, $-a^{-\lambda_{k+1}}\pi \leqq \theta \leqq a^{-\lambda_{\ +1}}\pi$,

$$\max | f_k(u_k) | \leqq Bc^{N_k} \exp\left\{- (Be)^{2^{N_k}}\right\}.$$

We may thus increase N_k if necessary so as to ensure that, on the same range,

$$\max | f_k(u_k) | \leqq 2^{-k}. \tag{9}$$

This finally fixes N_k.

4. We observe first that, in virtue of (6), $F(z)$ is in fact an integral function. Next (6) and (9) give us

$$\max\left\{ \sum_{l=1}^{k-1} + \sum_{l=k+1}^{\infty} | f_l(u_l) | \right\} \leqq B, \tag{10}$$

$$R_k \leqq | z | \leqq R_{k+1}, \; - a^{-\lambda_{k+1}}\pi \leqq \theta \leqq a^{-\lambda_{k+1}}\pi.$$

Also $F_{N_{k+1}}$ is so constructed that for fixed r, satisfying $R_k \leqq r \leqq R_{k+1}$,

$$\min | f_{k+1}(u_{k+1}) | \leqq B, \tag{11}$$

$$- a^{-\lambda_{k+1}}\pi \leqq \theta \leqq a^{-\lambda_{k+1}}\pi.$$

The equations (10) and (11) show that if r is fixed with $R_k \leqq r \leqq R_{k+1}$, then

$$\min_{|z|=r} | F(z) | \leqq B,$$

where B is independent of r, k. Thus, the minimum modulus of $F(z)$ on circles $|z| = r$ is bounded.

5. We have now to show that $F(z)$ has no asymptotic path. To do this we show that in certain regions the differential coefficient of $F(z)$ is not only large but so large that there can be no continuous path passing through all these regions on which $F(z)$ is bounded.

Consider $F'(z)$ in the annulus

$$1 - 3a^{-N_k} \leqq u_k \leqq 1 - 2a^{-N_k}. \tag{12}$$

We have

$$\frac{d}{dz}f_k(u_k) = \frac{\alpha_k}{z}u_k\frac{d}{du_k}f_k(u_k) = \frac{\alpha_k}{z}u_k(ac)^{N_k}\left(\frac{z}{R_k}\right)^{\beta_k-\alpha_k}(1+\epsilon),$$

where $|\epsilon| \leqq |10^{-k}$, in virtue of the remarks at the end of §2. Now, in the annulus considered, when $a>6$,

$$\left|\frac{\alpha_k}{z}u_k(ac)^{N_k}\left(\frac{z}{R_k}\right)^{\beta_k-\alpha_k}\right| = \beta_k c^{N_k}R_k^{-1}\left|\frac{z}{R_k}\right|^{\beta_k-1}$$
$$\geqq \beta_k c^{N_k}R_k^{-1}(1-3a^{-N_k})^{a^{N_k}} \geqq 10^{-2}\beta_k c^{N_k}R_k^{-1}.$$

Also, by (7) and (8), in the annulus considered,

$$\left|\frac{d}{dz}\left\{\sum_{m=1}^{k-1}+\sum_{m=k+1}^{\infty}f_m(u_m)\right\}\right| \leqq 10^{-6}\beta_k c^{N_k}R_k^{-1}+10^{-6},$$

and thus

$$F'(z) = \beta_k c^{N_k}R_k^{-\beta_k}z^{\beta_k-1}(1+\epsilon'), \tag{13}$$

where $|\epsilon'| \leqq 3\cdot 10^{-4} \leqq 10^{-3}\pi^{-1}$.

Now let ζ and ζ' be two points of the annulus (12) and let

$$\left|\frac{\zeta'}{\zeta}-1\right| = 10^{-1}\beta_k^{-1}.$$

Then (13) shows that

$$\begin{aligned} F(\zeta')-F(\zeta) &= \int_{\zeta'}^{\zeta}F'(z)dz \\ &= \int_{\zeta'}^{\zeta}\beta_k c^{N_k}R_k^{-\beta_k}z^{\beta_k-1}(1+\epsilon')dz \\ &= c^{N_k}\zeta^{\beta_k}R_k^{-\beta_k}\left\{\left(\frac{\zeta'}{\zeta}\right)^{\beta_k}-1\right\}+R, \end{aligned} \tag{14}$$

where

$$\begin{aligned} |R| &\leqq 10^{-3}|\zeta'-\zeta| \max_{|z|\leqq R_k}|\beta_k c^{N_k}R_k^{-\beta_k}z^{\beta_k-1}| \\ &\leqq 10^{-3}|\zeta'-\zeta|\beta_k c^{N_k}R_k^{-1} \leqq 10^{-4}c^{N_k}, \end{aligned} \tag{15}$$

for we may certainly find a path joining ζ and ζ' of length not exceeding $|\zeta'-\zeta|\pi$, entirely interior to the annulus (12). The first term of (14) is

$$c^{N_k}\zeta^{\beta_k}R_k^{-\beta_k}\{\beta_k(\zeta'-\zeta)/\zeta\}(1+\epsilon''), \tag{16}$$

where

$$|\epsilon''| \leqq 10\{(1 + 10^{-1}\beta_k^{-1})^{\beta_k} - 1 - 10^{-1}\} < 1/10.$$

Since, finally, in virtue of (12), if a is large enough,

$$|c^{N_k}\zeta^{\beta_k}R_k^{-\beta_k}\cdot\beta_k(\zeta' - \zeta)/\zeta| \geqq 10^{-3}c^{N_k},$$

it follows from (14), (15), (16) that

$$|F(\zeta') - F(\zeta)| \geqq 2\cdot 10^{-4}c^{N_k}. \tag{17}$$

Since, for sufficiently large a, the breadth of the strip (12) exceeds

$$10^{-1}R_k\beta_k^{-1},$$

(17) shows that there can be no continuous path, crossing the strip, for which the minimum modulus of $F(z)$ is less than $10^{-4}c^{N_k}$ which is arbitrarily large with k. Thus there can be no asymptotic path tending to infinity.

Massachusetts Institute of Technology,
Cambridge, Mass.

PARAMETRIZATIONS OF SADDLE SURFACES, WITH APPLICATION TO THE PROBLEM OF PLATEAU†

BY

E. J. McSHANE‡

Introduction. In the study of the properties of rectifiable curves $x=x(t)$, $y=y(t)$, $z=z(t)$, and of integrals of the calculus of variations taken along such curves, the investigator is greatly aided by the fact that the absolute continuity of $x(t)$, $y(t)$ and $z(t)$ is known to be necessary and sufficient in order that the length of the curve be equal to the classical integral $\int[x'^2+y'^2+z'^2]^{1/2}dt$, and also by the existence of a parametric representation of the curve (in terms of length of arc) in which the defining functions are Lipschitzian. On the other hand, let us suppose that the continuous surface S, represented by the equations $x=x(u, v)$, $y=y(u, v)$, $z=z(u, v)$, has finite area in the sense of Lebesgue. We know no conditions necessary and sufficient to insure that the area of S be equal to $\int\int(EG-F^2)^{1/2}\,dudv$, nor can we in general find a parametric representation of S which enjoys any particularly desirable properties. However, in a previous paper§ I have found certain conditions on the functions $x(u, v)$, $y(u, v)$, $z(u, v)$ which are sufficient to insure that the area be given by the classical double integral; and I have shown that on the class of all surfaces satisfying these conditions the double integrals of the kind usually considered in the calculus of variations have the property of semi-continuity.

It is therefore desirable to show that large classes of surfaces can be given representations satisfying the above mentioned conditions. Certainly it is not true that all continuous surfaces can be given such representations. However, let us restrict our attention to the class of surfaces for which the defining functions $x(u, v)$, etc., are monotonic in the sense of Lebesgue (including in particular the important class of saddle surfaces‖). In the present paper it is shown that for every such surface a representation can be found which satisfies the conditions mentioned, and is in fact almost as advantageous as a Lipschitzian representation would be.

---

† Presented to the Society, October 29, 1932; received by the editors July 19, 1932, and, after revision, December 20, 1932.

‡ National Research Fellow.

§ *Integrals over surfaces in parametric form*, Annals of Mathematics, vol. 34.

‖ The definition of the term "saddle surface" is given in §1.

One might readily suspect that such a representation would not be devoid of utility. As a matter of fact, a first use presents itself almost immediately; for with little additional effort we arrive at a solution† of the problem of Plateau, a solution not without interest when viewed as an application of general theorems proper to the direct method of the calculus of variations. To those readers who are interested principally in the problem of Plateau and only secondarily in the general theorems, I would like to point out that the distinctive feature of the present method is not its elegance (in which respect it is inferior to its predecessors) but the directness of the line of thought. First a solution of the problem of least area is found; then this solution is shown to admit of a representation which is in a sense almost everywhere conformal, so that the surface has to be minimal.

Finally, I would like to remark that whatever familiarity I may have with this branch of mathematics is due in large part to my conversations with Professor Radó.

1. Monotonic functions and saddle surfaces. Let the function $f(u, v)$ be defined and continuous on a point set E consisting of an open set plus its boundary. We define‡ the *monotonic deficiency* of f (as Lebesgue did) in the following manner. Let R be any open set in E and R^* its boundary, and denote the maximum and minimum of f on $R+R^*$ by L, l respectively, and the maximum and minimum of f on R^* by L_1, l_1 respectively. Then the least upper bound of the quantities $L-L_1$ and l_1-l as R varies over all open sets contained in E will be called the *monotonic deficiency* of f. Clearly this quantity is $\geqq 0$.

The function f will be called *monotonic* if its monotonic deficiency is zero.

Suppose now that S is a continuous surface§ represented by the equations

$$x = x(u, v),\ y = y(u, v),\ z = z(u, v),\ (u, v) \text{ on } B,$$

where B is a region consisting of a Jordan curve and its interior. The surface S is said to be a *saddle surface* provided that for every triple of constants a, b, c the linear combination $ax(u, v)+by(u, v)+cz(u, v)$ is a monotonic func-

† Previous solutions of this problem have been obtained by J. Douglas and by T. Radó. The solutions by these authors of the problem of Plateau in its usual form are summed up in the following papers:

J. Douglas, *Solution of the problem of Plateau*, these Transactions, vol. 33 (1931), p. 263;

T. Radó, *The problem of the least area and the problem of Plateau*, Mathematische Zeitschrift, vol. 32 (1930), p. 763.

‡ H. Lebesgue, *Sur le problème de Dirichlet*, Rendiconti del Circolo Matematico di Palermo, vol. 24 (1907), p. 385.

§ We assume that the reader is familiar with the definition of distance of continuous surfaces and of Lebesgue area, as presented, e.g., by Radó (loc. cit., pp. 772–774) or McShane (Annals of Mathematics, vol. 33, pp. 461–463).

tion. If S has everywhere a well-defined total curvature, this is equivalent to requiring that the total curvature be non-positive; if $x=u$ and $y=v$, so that $z=z(x, y)$, this is equivalent to Radó's definition.†

It is however our duty to show that the property of being a saddle surface is a property of the surface S, and does not depend on the particular representation of S. To do this we show that for every surface S it is true that *for every pair of representations*

$$(1.1) \qquad x = x(u, v), \quad y = y(u, v), \quad z = z(u, v), \quad (u, v) \text{ on } B,$$

$$(1.2) \qquad x = \bar{x}(\bar{u}, \bar{v}), \quad y = \bar{y}(\bar{u}, \bar{v}), \quad z = \bar{z}(\bar{u}, \bar{v}), \quad (\bar{u}, \bar{v}) \text{ on } \bar{B},$$

of the same surface S, and for every triple of constants a, b, c, the monotonic deficiency of $ax+by+cz$ is equal to the monotonic deficiency of $a\bar{x}+b\bar{y}+c\bar{z}$. The proof of this statement offers no difficulty, and from it our assertion concerning saddle surfaces follows immediately.

However, the class of surfaces for which we shall obtain a special representation is larger than the class of saddle surfaces, and consists in fact of those surfaces for which the defining functions $x(u, v)$, $y(u, v)$, $z(u, v)$ are each monotonic. The point particularly to be observed is that this property is independent of the particular representation of the surface.

2. **Lemma on convergence.** The functions with which we shall be particularly concerned are those satisfying the following conditions.

(2.1a) The function $f(u, v)$ is defined and continuous on the unit circle C: $u^2+v^2 \leqq 1$.

(2.1b) $f(u, v)$ is absolutely continuous in u for almost all fixed values of v, and absolutely continuous in v for almost all fixed values of u.

(2.1c) The integral

$$I(f) \equiv \iint_{u^2+v^2<1} [(\partial f/\partial u)^2 + (\partial f/\partial v)^2]\, du\, dv$$

exists.‡

We can now state an extension of a theorem of Lebesgue§ which is to play an important role in the following pages.

† See, e.g., Radó, *Geometrische Betrachtungen über zweidimensionale reguläre Variationsprobleme*, Acta Szeged, vol. 2 (1926), pp. 228–253, especially pp. 229–230.

‡ This and all other integrals are understood to be Lebesgue integrals.

§ H. Lebesgue, loc. cit., p. 386.

LEMMA 1. *Let the functions* $f_1(u, v), f_2(u, v), \cdots$ *be defined and satisfy conditions* (2.1) *on the unit circle* C. *Suppose further that the sequence* $\{f_n(u, v)\}$ *is uniformly convergent on the circumference of* C, *and that the monotonic deficiency of* $f_n(u, v)$ *tends to zero with* $1/n$, *and also that* $I(f_n) \leqq H$ *for every* n, *where* H *is some constant. Then there exists a subsequence of the* $\{f_n(u, v)\}$ *converging uniformly on the whole circle* C *to a monotonic limit function* $f(u, v)$, *and* $f(u, v)$ *also satisfies conditions* (2.1).

To prove the existence of a uniformly convergent subsequence of $\{f_n\}$ we can follow the proof of Lebesgue† almost without change. The fact that the $f_n(u, v)$ are not all identical on the boundary, but merely converge uniformly, causes no trouble. But we find it convenient to replace the circles $C(r)$ used by Lebesgue in his proof by squares $Q(r)$; this device will be used in the proof of Lemma 2, and we therefore do not give it in detail here.

It remains to prove that the limit function $f(u, v)$ satisfies conditions (2 1). To do this we use a slight modification of a theorem of Fatou‡: *If the functions* $\phi_1(x), \phi_2(x), \cdots$ *are all non-negative and their integrals*

$$\int_a^b \phi_n(x)dx$$

are all less than a fixed number, then lim inf $\phi_n(x)$ *is summable, and*

$$\int_a^b \liminf \phi_n(x)dx \leqq \liminf \int_a^b \phi_n(x)dx.$$

The proof differs only very slightly from that given by Fatou. Returning to our uniformly convergent subsequence of the functions $f_n(u, v)$ (for which subsequence we retain the notation $\{f_n(u, v)\}$), we find by virtue of this theorem that§

$$\int_{-1}^{1} du \cdot \liminf \int dv(\partial f_n/\partial v)^2 \leqq \liminf \int_{-1}^{1} du \cdot \int dv(\partial f_n/\partial v)^2 \leqq H.$$

Hence the set of values of u for which the expression

$$\liminf \int (\partial f_n/\partial v)^2 dv \tag{2.2}$$

† Loc. cit.

‡ P. Fatou, *Séries trigonométriques et séries de Taylor*, Acta Mathematica, vol. 30 (1906), p. 375.

§ All single integrals in this proof are understood to have the limits $-(1-u^2)^{1/2}, +(1-u^2)^{1/2}$, unless other ranges are specifically indicated.

is infinite has measure zero. Likewise, since $(\partial f_n/\partial u)^2$ is summable over the unit circle, the integral

$$\int (\partial f_n/\partial v)^2 dv \tag{2.3}$$

is finite for almost all values of u. We define E to be the set of values of u for which one or more of the expressions (2.2), (2.3) is infinite; this set has measure zero, and we shall henceforward restrict our attention to values of u belonging to the complement of E.

We now fix upon any value of u belonging to the complement of E, and select a subsequence $f_\alpha(u, v)$ of our original convergent subsequence (α ranges over a subset of the positive integers) for which there exists the limit

$$\lim_{\alpha\to\infty} \int (\partial f_\alpha(u, v)/\partial v)^2 dv = \liminf \int (\partial f_n(u, v)/\partial v)^2 dv.$$

By the proof of a theorem of F. Riesz† it is possible to select a subsequence of the sequence $\{f_\alpha\}$ which converges everywhere to a function $\phi(v)$ which is absolutely continuous and whose derivative $d\phi/dv$ is summable together with its square, and in addition

$$\int (d\phi/dv)^2 dv \leqq \lim_{\alpha\to\infty} \int (\partial f_\alpha/\partial v)^2 dv.$$

But the sequence $\{f_n\}$ converges uniformly to $f(u, v)$; hence $\phi(v) = f(u, v)$, so that $f(u, v)$ is absolutely continuous in v and

$$\int (\partial f/\partial v)^2 dv \leqq \liminf \int (\partial f_n/\partial v)^2 dv.$$

Applying Fatou's theorem once again, we obtain

$$\begin{aligned} \int_{CE} du \int dv (\partial f/\partial v)^2 &\leqq \int_{CE} du \cdot \liminf \int dv (\partial f_n/\partial v)^2 \\ &\leqq \liminf \iint_{u^2+v^2<1} (\partial f_n/\partial v)^2 du dv \leqq H. \end{aligned} \tag{2.4}$$

Likewise $f(u, v)$ is absolutely continuous in u for almost all fixed values of v, and

$$\iint_{u^2+v^2<1} (\partial f/\partial u)^2 du dv \leqq \liminf \iint_{u^2+v^2<1} (\partial f_n/\partial u)^2 du dv \leqq H. \tag{2.5}$$

† F. Riesz, *Untersuchungen über Systeme integrierbarer Funktionen*, Mathematische Annalen, vol. 69 (1910), pp. 466–468.

Adding inequalities (2.4) and (2.5), we† find that $I(f)$ exists and moreover

$$I(f) \leqq \liminf I(f_n). \tag{2.6}$$

3. Convergence on the boundary. For the sake of compactness, the boundary curve

$$x = x(\cos\theta, \sin\theta),\ y = y(\cos\theta, \sin\theta),\ z = z(\cos\theta, \sin\theta)$$

of the surface

$$x = x(u, v),\ y = y(u, v),\ z = z(u, v),\ u^2 + v^2 \leqq 1,$$

will be written in the shorter form

$$x = x(\theta),\ y = y(\theta),\ z = z(\theta).$$

The symbols E, F, G will as usual have the respective meanings $x_u^2 + y_u^2 + z_u^2$, $x_u x_v + y_u y_v + z_u z_v$, $x_v^2 + y_v^2 + z_v^2$.

We now proceed to prove our second lemma.

Lemma 2. *Let* $\{S_n\}$ *be a sequence of surfaces possessing representations*

$$x = x_n(u, v),\ y = y_n(u, v),\ z = z_n(u, v),\ u^2 + v^2 \leqq 1, \tag{3.1}$$

in which the functions x_n, y_n, z_n *satisfy conditions* (2.1), *and for which*

$$\iint_{u^2+v^2<1} (E_n + G_n)\,du\,dv \leqq H$$

for all values of n, H *being a fixed number. Suppose further that the boundary curves*

$$\Gamma_n\colon\ x = x_n(\theta),\ y = y_n(\theta),\ z = z_n(\theta) \tag{3.2}$$

of the surfaces S_n *approach as a limit a simple closed curve* Γ, *and that for three distinct values* θ_1, θ_2, θ_3 *of* θ *the sequences of points* $\{x_n(\theta_i), y_n(\theta_i), z_n(\theta_i)\}$ *approach three distinct limit points* (ξ_i, η_i, ζ_i) $(i = 1, 2, 3)$. *Then it is possible to find a representation*

$$x = x(\theta),\ y = y(\theta),\ z = z(\theta) \tag{3.3}$$

of the curve Γ *and a subsequence of the* $\{S_n\}$ *(for which we retain the same notation) such that*

$$\lim x_n(\theta) = x(\theta), \quad \lim y_n(\theta) = y(\theta), \quad \lim z_n(\theta) = z(\theta)$$

uniformly in θ.

---

† My original proof of this lemma was decidedly more intricate than the above; for the present proof I wish here to thank Professor Tamarkin.

Let $\{\epsilon_n\}$ be a sequence of positive numbers tending to zero such that the distance $\|\Gamma_n, \Gamma\|$ between Γ_n and Γ is less than ϵ_n for every n. If we fix upon any topological representation $x=x(\tau)$, $y=y(\tau)$, $z=z(\tau)$ of the curve Γ on the unit circumference (the functions x, y, z having period 2π), then we can find a topological mapping of the unit circumference on itself, expressible by the equation $\tau=\tau_n(\theta)$, $0\leqq\theta\leqq 2\pi$, in which $\tau_n(\theta)$ is a continuous monotonic function such that $0\leqq\tau(0)=\tau_n(2\pi)-2\pi\leqq 2\pi$, for which

$$(3.4)\qquad | x_n(\theta) - x(\tau_n(\theta)) | < \epsilon_n,$$

with like inequalities for y and z. By Helly's theorem there exists a subsequence of the sequence $\{\tau_n(\theta)\}$ which converges for each θ in the interval $(0, 2\pi)$ to a monotonic limit function $\tau(\theta)$. We reject all τ_n and their corresponding S_n which do not belong to this subsequence, and we re-name the remaining subsequence $\{\tau_n\}$, and the remaining surfaces $\{S_n\}$.

We thus find that for every θ we have $\lim \tau_n(\theta)=\tau(\theta)$; and since $x(\tau)$, $y(\tau)$, $z(\tau)$ are continuous, this implies $\lim x(\tau_n(\theta))=x(\tau(\theta))$ for every θ. From this and inequality (3.4) it follows that for every θ we have

$$(3.5)\qquad \lim x_n(\theta) = x(\tau(\theta));$$

similar equations hold for y and z. In particular, for the values θ_1, θ_2, θ_3 of the hypothesis we have $\lim x_n(\theta_i)=\xi_i$, whence $x(\tau(\theta_i))=\xi_i$; likewise $y(\tau(\theta_i)) =\eta_i$, $z(\tau(\theta_i))=\zeta_i$. Since the three points (ξ_i, η_i, ζ_i) are distinct, the three numbers $\tau(\theta_i)$ must also be distinct.

We now prove that the function $\tau(\theta)$ is continuous. Suppose on the contrary that there is a point† θ_0 of discontinuity of $\tau(\theta)$; then $\tau(\theta_0-0)$ and $\tau(\theta_0+0)$ both exist, and $\tau(\theta_0+0)>\tau(\theta_0-0)$. We cannot have $\tau(\theta_0+0) =\tau(\theta_0-0)+2\pi$; for this would imply that $\tau(\theta)$ is constantly equal to $\tau(\theta_0-0)$ on the interval $(0, \theta_0)$, and constantly equal to $\tau(\theta_0+0)$ on the interval $(\theta_0, 2\pi)$, as readily follows from the fact that $\tau(\theta)$ is monotonic and $\tau(2\pi)=\tau(0)+2\pi$. This is in contradiction with the previously established fact that $\tau(\theta_1)$, $\tau(\theta_2)$, and $\tau(\theta_3)$ are all distinct. Hence $0<\tau(\theta_0+0)-\tau(\theta_0-0) <2\pi$; and since the equations $x=x(\tau)$, etc., map the unit circumference topologically on Γ, the points

$$(x(\tau(\theta_0 - 0)),\ y(\tau(\theta_0 - 0)),\ z(\tau(\theta_0 - 0)))$$

and

$$(x(\tau(\theta_0 + 0)),\ y(\tau(\theta_0 + 0)),\ z(\tau(\theta_0 + 0)))$$

are distinct. Suppose to be specific that

† The following proof is constructed for the case in which θ_0 is an interior point of the interval $(0, 2\pi)$. In case θ_0 is an end point of the interval, say $\theta_0=0$, we have only to remember that $\tau(\theta)$ and all the $\tau_n(\theta)$ are periodic, and consider them as defined on the interval $(-\pi, \pi)$.

$$x(\tau(\theta_0 + 0)) = x(\tau(\theta_0 - 0)) + 3\epsilon, \epsilon > 0. \tag{3.6}$$

Since $x(\tau)$ is continuous, there exists a positive number δ ($<\pi/3$) such that

$$| x(\tau(\theta)) - x(\tau(\theta_0 - 0)) | < \epsilon \qquad (\theta_0 - \delta < \theta < \theta_0) \tag{3.7}$$

and

$$| x(\tau(\theta)) - x(\tau(\theta_0 + 0)) | < \epsilon \qquad (\theta_0 < \theta < \theta_0 + \delta). \tag{3.8}$$

We now introduce a new coordinate system convenient for our present purposes. Let $Q(r)$ be the square (the line-configuration, not the region) with center at the point ($\cos \theta_0$, $\sin \theta_0$) and sides of length $2r$ parallel to the coordinate axes. If $4r \leqq \delta$, the square $Q(r)$ intersects the circumference once in the interval $(\theta_0 - \delta, \theta_0)$ and once in the interval $(\theta_0, \theta_0 + \delta)$. To each point (u, v) on $Q(r)$ we assign the coordinates (r, s), where s is the length of arc of $Q(r)$ from the point at which $Q(r)$ enters the circle C to the point (u, v), measured counter-clockwise. Then $(\partial x_n(r, s)/\partial s)^2$ is equal to $(\partial x_n(u, v)/\partial u)^2$ or to $(\partial x_n(u, v)/\partial v)^2$, according to the side of $Q(r)$ on which (u, v) lies; and therefore denoting by $s(r)$ the s-coordinate of the point at which $Q(r)$ leaves the circle, we have

$$\begin{aligned} &\int_0^{\delta/4}\int_0^{s(r)} (\partial x_n(r, s)/\partial s)^2 ds dr \\ &\qquad \leqq \iint_{u^2+v^2<1} [(\partial x_n/\partial u)^2 + (\partial x_n/\partial v)^2] du dv \leqq H. \end{aligned} \tag{3.9}$$

Hence for almost all values of $r \leqq \delta/4$ all the integrals

$$\int_0^{s(r)} (\partial x_n/\partial s)^2 ds \tag{3.10}$$

are finite; and by the theorem of Fatou†

$$\int_0^{\delta/4} \liminf \int_0^{s(r)} (\partial x_n/\partial s)^2 ds \leqq \liminf \int_0^{\delta/4}\int_0^{s(r)} (\partial x_n/\partial s)^2 ds dr \leqq H \tag{3.11}$$

so that for almost all values of $r \leqq \delta/4$ the expression

$$\liminf \int_0^{s(r)} (\partial x_n/\partial s)^2 ds \tag{3.12}$$

is finite. We define E to be the set of values of r for which one or more of the expressions (3.10) or (3.12) is infinite; then E has measure zero. We henceforth consider only values of r which belong to the complement of E, and

† I.e., as stated in the proof of Lemma 1.

which satisfy the additional condition that all the functions $x_n(r, s)$ are absolutely continuous in s; the set of values of r thus rejected has measure zero.

To the integrals (3.10) we now apply the inequality of Schwarz, thus finding

$$\begin{aligned}\int_0^{s(r)} (\partial x_n/\partial s)^2 ds &\geqq \left\{ \int_0^{s(r)} (\partial x_n/\partial s) ds \right\}^2 \cdot \frac{1}{s(r)} \\ &= [x_n(r, s(r)) - x_n(r, 0)]^2/s(r).\end{aligned} \tag{3.13}$$

Inequalities (3.13) and (3.11) together imply that

$$\int_0^{\delta/4} \liminf\, [x_n(r, s(r)) - x_n(r, 0)]^2 \cdot \frac{1}{s(r)} \cdot dr \leqq H. \tag{3.14}$$

But the points $(r, 0)$ and $(r, s(r))$ are both on the circumference of the circle C, and lie each in one of the intervals $(\theta_0 - \delta, \theta_0)$ and $(\theta_0, \theta_0 + \delta)$; hence by (3.6), (3.7), and (3.8) we have

$$\liminf\, [x_n(r, s(r)) - x_n(r, 0)]^2 \geqq \epsilon^2. \tag{3.15}$$

Also by its definition $s(r) < 8r$. Hence

$$\liminf\, [x_n(r, s(r)) - x_n(r, 0)]^2 \cdot \frac{1}{s(r)} > \frac{\epsilon^2}{8r}, \tag{3.16}$$

which is not summable over the interval $0 \leqq r \leqq \delta/4$. This contradicts inequality (3.14), and therefore the assumption that $\tau(\theta)$ is discontinuous leads to a contradiction.

Therefore the sequence $\{\tau_n(\theta)\}$ of continuous monotonic functions converges everywhere to the continuous function $\tau(\theta)$. It follows† that the convergence of $\tau_n(\theta)$ to $\tau(\theta)$ is uniform; hence $\lim x(\tau_n(\theta)) = x(\tau(\theta))$ uniformly. This with (3.4) implies that $\lim x_n(\theta) = x(\tau(\theta))$ uniformly in θ; similar statements hold for y, z. Hence the curve

$$x = x(\tau(\theta)),\ y = y(\tau(\theta)),\ z = z(\tau(\theta)) \tag{3.17}$$

is a limit curve of the sequence $\{\Gamma_n\}$. But $\{\Gamma_n\}$ has the unique limit Γ; therefore equations (3.17) form a representation of Γ, and the lemma is established.

4. **A theorem on representations.** We now proceed to prove our principal theorem concerning parametric representations.

---

† Buchanan, H. E., and Hildebrandt, T. H., *Note on the convergence of a sequence of functions of a certain type*, Annals of Mathematics, vol. 9 (1908), p. 123.

THEOREM I. *If the continuous surface S, represented by the equations*

$$x = \bar{x}(u, v),\ y = \bar{y}(u, v),\ z = \bar{z}(u, v),\ u^2 + v^2 \leqq 1,$$

satisfies the conditions

(4.1) *the Lebesgue area $L(S)$ of the surface S is finite;*

(4.2) *the curve*

$$\Gamma:\ x = \bar{x}(\theta),\ y = \bar{y}(\theta),\ z = \bar{z}(\theta)$$

bounding S is a Jordan curve;

(4.3) *the functions $\bar{x}(u, v)$, $\bar{y}(u, v)$, $\bar{z}(u, v)$ are monotonic;*

then there exists a representation

$$x = x(u, v),\ y = y(u, v),\ z = z(u, v),\ u^2 + v^2 \leqq 1,$$

of S in which the functions $x(u, v)$, etc., satisfy the conditions (2.1), *and also satisfy the relations*

$$E = G,\ F = 0 \textit{ for almost all values of } (u, v). \tag{4.4}$$

Moreover under any change of parameters $u=u(\bar{u}, \bar{v})$, $v=v(\bar{u}, \bar{v})$ representing a conformal mapping of the unit circle on itself the three functions $x(u(\bar{u}, \bar{v}), v(\bar{u}, \bar{v}))$, etc., also satisfy conditions (2.1) *and* (4.4).

Before proceeding to the proof of this theorem, we observe that the hypotheses are independent of the representation of S.

Now let $\{\Pi_n\}$ be a sequence of polyhedra tending to S for which the areas $L(\Pi_n)$ tend to $L(S)$; we can assume without loss of generality that none of the triangles which form the faces of Π_n are degenerate. If Π_n has the representation

$$x = \bar{x}_n(\alpha, \beta),\ y = \bar{y}_n(\alpha, \beta),\ z = \bar{z}_n(\alpha, \beta),\ \alpha^2 + \beta^2 \leqq 1,$$

then there exists for each n a topological mapping $\alpha=\alpha_n(u, v)$, $\beta=\beta_n(u, v)$ of the unit circle on itself such that $\lim \bar{x}_n(\alpha_n(u, v), \beta_n(u, v)) = \bar{x}(u, v)$ uniformly; and since $\bar{x}(u, v)$ is monotonic, this implies† that the monotonic deficiency of $\bar{x}_n(\alpha_n(u, v), \beta_n(u, v))$ tends to zero with $1/n$; similar statements hold for y and for z. But the functions $\bar{x}_n(\alpha_n(u, v), \beta_n(u, v))$, etc., form a representation of Π_n; hence by §1 the monotonic deficiency of $\bar{x}_n(\alpha, \beta)$ is equal to the monotonic deficiency of $\bar{x}_n(\alpha_n(u, v), \beta_n(u, v))$, and consequently tends to zero with $1/n$.

On the circumference of the unit circle we now choose three distinct points A_1^*, A_2^*, A_3^*, and on the curve we choose three distinct points A_1, A_2, A_3.

---

† Lebesgue, loc. cit., p. 385.

Since $\lim \Pi_n = S$, the boundary curves Γ_n of the polyhedra Π_n tend to the curve Γ; hence on each Γ_n we can choose three distinct points $A_1^{(n)}, A_2^{(n)}, A_3^{(n)}$ such that $A_i^{(n)}$ approaches A_i as n increases.

From general theorems on the conformal maps of abstract Riemann surfaces† it follows that any polyhedron Π whose faces are non-degenerate admits of a parametric representation of the following kind.

(a) The functions representing Π are defined on the unit circle; i.e., Π is given by the equations

$$x = x(u, v),\ y = y(u, v),\ z = z(u, v),\ u^2 + v^2 \leqq 1. \tag{4.5}$$

(b) The unit circle is subdivided by arcs into a finite number of curvilinear triangles $\delta_1, \cdots, \delta_k$, and the equations (4.5) carry each δ_i in a topological way into a rectilinear triangle in xyz-space.

(c) The triangles δ_i are bounded by arcs which are analytic, including end points.

(d) Interior to each triangle δ_i the functions $x(u, v)$, $y(u, v)$, $z(u, v)$ are analytic and satisfy the relations

$$E = G, F = 0. \tag{4.6}$$

(e) Three arbitrarily given distinct points A_1, A_2, A_3 on the boundary curve of Π correspond under equations (4.5) to three arbitrarily given distinct points A_1^*, A_2^*, A_3^* on the unit circle $u^2+v^2=1$.

For such a representation of Π we find without difficulty that conditions (2.1) are satisfied. We now choose for each Π_n a representation

$$\Pi_n:\ x = x_n(u, v),\ y = y_n(u, v),\ z = z_n(u, v),\ u^2 + v^2 \leqq 1, \tag{4.7}$$

satisfying the above conditions; in particular, for the points A_1^*, A_2^*, A_3^* we choose the points already so named on the circumference of the unit circle, and for the points A_1, A_2, A_3 we choose the points $A_1^{(n)}, A_2^{(n)}, A_3^{(n)}$.

We now make use of the theorem‡ that *if the functions* $x(u, v)$, $y(u, v)$, $z(u, v)$, $u^2+v^2 \leqq 1$, *satisfy conditions* (2.1), *then the area of the surface* S: $x=x(u, v)$, $y=y(u, v)$, $z=z(u, v)$, $u^2+v^2\leqq 1$, *is equal to*

---

† See, for instance, Carathéodory, *Conformal Representation* (No. 28 of the Cambridge Tracts in Mathematics and Mathematical Physics), in particular chapter VII. While the theorem on the conformal maps of polyhedra can be obtained as a special case of general facts, it should be observed that this theorem was proved by H. A. Schwarz.

‡ E. J. McShane, loc. cit. in introduction. This theorem has also been established independently and by different methods by C. B. Morrey, in a paper not as yet published.

This theorem is needed later, but for the present case it is stronger than necessary; equation (4.8) can be established by simpler means. Cf. Radó, loc. cit., p. 774.

$$\iint_{u^2+v^2<1} (EG - F^2)^{1/2} du dv.$$

Applying this to the polyhedra Π_n, we find

$$L(\Pi_n) = \iint_{u^2+v^2<1} (E_n G_n - F_n^2)^{1/2} du dv; \tag{4.8}$$

and by (4.6) this implies

$$L(\Pi_n) = \iint_{u^2+v^2<1} \tfrac{1}{2}(E_n + G_n) du dv. \tag{4.9}$$

Since $\lim L(\Pi_n) = L(S)$, the right member of (4.9) is bounded. Observing that the equation $\lim \Pi_n = S$ implies that the boundary curves Γ_n of the polyhedra Π_n converge to Γ, we find that all of the hypotheses of Lemma 2 are satisfied. We can therefore select a representation $x = x(\theta)$, $y = y(\theta)$, $z = z(\theta)$ of Γ and a subsequence of the $\{\Pi_n\}$ (for which we retain the same notation) such that

$$\lim x_n(\theta) = x(\theta), \quad \lim y_n(\theta) = y(\theta), \quad \lim z_n(\theta) = z(\theta) \tag{4.10}$$

uniformly in θ.

This subsequence now satisfies the hypotheses of Lemma 1, and we can therefore select from it a subsequence (which we continue to call $\{\Pi_n\}$) for which the functions $x_n(u, v)$, $y_n(u, v)$, $z_n(u, v)$ converge uniformly on the whole circle to limit functions $x(u, v)$, $y(u, v)$, $z(u, v)$ which satisfy conditions (2.1). Therefore the surface defined by the equations

$$x = x(u, v), \quad y = y(u, v), \quad z = z(u, v), \quad u^2 + v^2 \leqq 1, \tag{4.11}$$

is a limit surface of the sequence $\{\Pi_n\}$. But the sequence has the unique limit S; hence equations (4.11) form a new representation of the surface S.

Since the functions $x(u, v)$, $y(u, v)$, $z(u, v)$ satisfy conditions (2.1), the area is given by the classical integral; hence

$$\begin{aligned} \iint_{u^2+v^2<1} \tfrac{1}{2}(E + G) du dv &\geqq \iint_{u^2+v^2<1} (EG)^{1/2} du dv \\ &\geqq \iint_{u^2+v^2<1} (EG - F^2)^{1/2} du dv = L(S). \end{aligned} \tag{4.12}$$

On the other hand, we know by (4.9) and (2.6) that

$$\lim L(\Pi_n) = \lim \iint_{u^2+v^2<1} \tfrac{1}{2}(E_n + G_n) du dv \geqq \iint_{u^2+v^2<1} \tfrac{1}{2}(E + G) du dv; \tag{4.13}$$

and since $\lim L(\Pi_n) = L(S)$, these inequalities imply

$$(4.14)\quad L(S) = \iint_{u^2+v^2<1} \tfrac{1}{2}(E+G)dudv = \iint_{u^2+v^2<1} (EG-F^2)^{1/2}dudv.$$

Now $(E+G)/2$ is never less than $(EG-F^2)^{1/2}$, and if E, F and G are finite the two can be equal only if $E=G$ and $F=0$. Hence from (4.14) it follows that

$$(4.15)\qquad E = G,\ F = 0 \text{ for almost all values of } (u, v).$$

It remains only to show that these properties ((2.1) and (4.15)) of $x(u, v)$, $y(u, v)$, $z(u, v)$ remain invariant under conformal mappings

$$(4.16)\qquad u = u(\bar{u}, \bar{v}),\ v = v(\bar{u}, \bar{v})$$

of the unit circle on itself. We retain the sequence of polyhedra $\{\Pi_n\}$ obtained above for which the representing functions $x_n(u, v)$, $y_n(u, v)$, $z_n(u, v)$ are uniformly convergent. Apply to each of these the transformation (4.16); we obtain a new representation

$$(4.17)\quad x = \bar{x}_n(\bar{u}, \bar{v}) \equiv x_n(u(\bar{u}, \bar{v}), v(\bar{u}, \bar{v})),\ y = \bar{y}_n(\bar{u}, \bar{v}),\ z = \bar{z}_n(\bar{u}, \bar{v}).$$

Since the mapping (4.16) is conformal, the functions $\bar{x}_n(\bar{u}, \bar{v})$, etc., continue to satisfy all the conditions (a), $\cdots$, (e) stated above. Also the sequences $\bar{x}_n(\bar{u}, \bar{v})$, etc., are uniformly convergent; let the limit function be $\bar{x}(\bar{u}, \bar{v})$. Then by the definition (4.17) of $\bar{x}_n(\bar{u}, \bar{v})$, we have $\bar{x}(\bar{u}, \bar{v}) = x(u(\bar{u}, \bar{v}), v(\bar{u}, \bar{v}))$, with like equations for y and z. All the arguments of the preceding paragraph are applicable, and we thus find that $\bar{x}(\bar{u}, \bar{v})$, etc., also satisfy conditions (2.1) and (4.4). The theorem is thus established.

5. **An area-reducing alteration.** We shall in studying the problem of Plateau have need of one further lemma.

LEMMA 3. *Let the continuous surface*

$$(5.1)\qquad S:\ x = x(u, v),\ y = y(u, v),\ z = z(u, v),\ u^2 + v^2 \leqq 1,$$

have finite area. Then there exists a surface $\bar{S}$: $x = \bar{x}(u, v)$, $y = \bar{y}(u, v)$, $z = \bar{z}(u, v)$, $u^2+v^2 \leqq 1$, *having the same boundary curve as* S, *and satisfying the conditions*

$$(5.2)\qquad L(\bar{S}) \leqq L(S),$$

(5.3) *the functions* $\bar{x}(u, v)$, $\bar{y}(u, v)$, $\bar{z}(u, v)$ *are monotonic.*

We need only a slight modification of a proof due to Lebesgue.† Let us arrange the rational numbers in a sequence r_1, r_2, $\cdots$. The point set at which $x(u, v) > r_1$ is (if not null) an open‡ set, and consists of a finite or denumerable set of maximal open connected sets. We disregard those sets

† Loc. cit., p. 382.

‡ Except that it may contain limit points on the circumference.

which have points in common with the circumference $u^2+v^2=1$, and name the remainder $R_1, R_2, \cdots$. We treat similarly the point set for which $x(u, v)<r_1$; the maximal connected open point sets, interior to the circle, on which $x(u, v)<r_1$ we call $T_1, T_2, \cdots$. We designate by $x^{(1)}(u, v)$ the function equal to r_1 on $R_1+T_1+R_2+T_2+\cdots$, and equal to $x(u, v)$ elsewhere; this is a continuous function. Moreover, the area of the surface

$$(5.4)\qquad S^{(1)}:\quad x = x^{(1)}(u, v),\ y = y(u, v),\ z = z(u, v)$$

is at most equal to $L(S)$. This is obvious if S is a polyhedron, for then the images of R_1, T_1, etc., under (5.1) consist of a finite number of triangles, and the images under (5.4) are the projections of those triangles on the plane $x=r_1$. For the general case, we select (as is always possible) a sequence $\{\Pi_n\}$ of polyhedra represented in the form

$$(5.5)\qquad x = x_n(u, v),\ y = y_n(u, v),\ z = z_n(u, v),\ u^2 + v^2 \leqq 1,$$

such that $\lim L(\Pi_n)=L(S)$, and $\lim x_n(u, v)=x(u, v)$ uniformly on the circle, with similar statements for y and z. We obtain $x_n^{(1)}(u, v)$ from $x_n(u, v)$ as above, and denote by $\Pi_n^{(1)}$ the polyhedron

$$x = x_n^{(1)}(u, v),\ y = y_n(u, v),\ z = z_n(u, v),\ u^2 + v^2 \leqq 1.$$

Then $\lim \Pi_n^{(1)}=S^{(1)}$, and $L(\Pi_n^{(1)})\leqq L(\Pi_n)$; hence

$$L(S^{(1)}) \leqq \liminf L(\Pi_n^{(1)}) \leqq \lim L(\Pi_n) = L(S).$$

We now obtain $x^{(2)}(u, v)$ from $x^{(1)}(u, v)$ as we obtained $x^{(1)}(u, v)$ from $x(u, v)$, the number r_2 taking the place of r_1. For the corresponding surface $S^{(2)}$ we have $L(S^{(2)})\leqq L(S)$. Likewise we obtain $x^{(3)}$ from $x^{(2)}$, using r_3 instead of r_2; and so on. The sequence of functions $\{x^{(n)}(u, v)\}$ converges uniformly to a limit function $\bar{x}(u, v)$, which is continuous and monotonic; the proof of this is identical with that given by Lebesgue†, and we shall not repeat it. We wish to emphasize two points; first, the functions $x^{(n)}(u, v)$ are all equal to $x(u, v)$ on the circumference $u^2+v^2=1$, so that $\bar{x}(\theta)=x(\theta)$; and second, the surface S_x defined by the equations $x=\bar{x}(u, v)$, $y=y(u, v)$, $z=z(u, v)$ has area at most equal to $L(S)$, since

$$L(S_x) \leqq \liminf L(S^{(n)}) \leqq L(S).$$

The whole process being repeated for the function $y(u, v)$, we obtain a monotonic function $\bar{y}(u, v)$, equal to $y(u, v)$ on the circumference, and such that the surface defined by the equations $x=\bar{x}(u, v)$, $y=\bar{y}(u, v)$, $z=z(u, v)$ has area at most equal to $L(S)$. Finally we repeat the whole process for

† Loc. cit.; beginning at the middle of p. 382.

$z(u, v)$, and obtain a monotonic function $\bar{z}(u, v)$, equal to $z(u, v)$ on the circumference $u^2+v^2=1$, and such that the surface $\bar{S}$ defined by the equations

$$x = \bar{x}(u, v),\ y = \bar{y}(u, v),\ z = \bar{z}(u, v),\ u^2 + v^2 \leqq 1,$$

has area $L(\bar{S}) \leqq L(S)$. The lemma is thus established.

6. **The problem of Plateau.** Let Γ be any Jordan curve in xyz-space; we designate by $a(\Gamma)$ the greatest lower bound of the areas of all continuous surfaces S bounded by Γ. On the other hand, let $\{\Pi_n\}$ be a sequence of polyhedra whose boundary curves tend to Γ, and consider the quantity lim inf $L(\Pi_n)$. The greatest lower bound of this quantity for all such sequences $\{\Pi_n\}$ is called the minimum area of Γ; we denote it by $m(\Gamma)$. There is no essential restriction in assuming that each polyhedron Π_n of the sequence be bounded by a Jordan polygon; for given Π_n, we can alter it so as to make the boundary non-self-intersecting while changing the area and displacing the boundary by arbitrarily small amounts.

It is easy to show that

$$m(\Gamma) \leqq a(\Gamma). \tag{6.1}$$

For let S be any continuous surface bounded by Γ, and let $\{\Pi_n\}$ be a sequence of polyhedra tending to S and such that $L(\Pi_n)$ tends to $L(S)$. The boundaries Γ_n of the polyhedra Π_n tend to Γ, and therefore

$$m(\Gamma) \leqq \liminf L(\Pi_n) = L(S);$$

this being true for every surface S bounded by Γ, inequality (6.1) follows immediately. We shall later prove (as is already known†) that $a(\Gamma) = m(\Gamma)$ for every Jordan curve Γ.

We now proceed to prove

THEOREM‡ II. *For every Jordan curve Γ whose minimum area $m(\Gamma)$ is finite, there exists a continuous surface*

$$S\colon\ x = x(u, v),\ y = y(u, v),\ z = z(u, v),\ u^2 + v^2 \leqq 1, \tag{6.2}$$

bounded by Γ and satisfying the following conditions:

(a) *the area of S is the least possible among all surfaces bounded by Γ, i.e.,*

$$L(S) = a(\Gamma) = m(\Gamma); \tag{6.3}$$

(b) *the functions $x(u, v)$, $y(u, v)$, and $z(u, v)$ are analytic, and in fact harmonic, for $u^2+v^2<1$;*

(c) *the surface S is a minimal surface.*

† Radó, loc. cit., p. 776.

‡ Douglas, loc. cit.; Radó, loc. cit., p 791

Let $\{\Pi_n\}$ be a sequence of polyhedra having areas $L(\Pi_n)$ tending to $m(\Gamma)$, and bounded by Jordan polygons Γ_n tending to Γ. By Lemma 3, for each n we can find a surface S_n bounded by Γ_n and of area $L(S_n) \leqq L(\Pi_n)$ for which the representing functions $\bar{x}_n(u, v)$, $\bar{y}_n(u, v)$, $\bar{z}_n(u, v)$ are monotonic. Now let A_1^*, A_2^*, A_3^* be three distinct points on the unit circumference $u^2+v^2=1$, and let A_1, A_2, A_3 be three distinct points on Γ. Since Γ_n tends to Γ, we can on each Γ_n select three points $A_1^{(n)}$, $A_2^{(n)}$, $A_3^{(n)}$ such that $A_i^{(n)}$ tends to A_i $(i=1, 2, 3)$. By Theorem I there exists a representation

$$x = x_n(u, v),\ y = y_n(u, v),\ z = z_n(u, v),\ u^2 + v^2 \leqq 1, \tag{6.4}$$

of S_n such that the functions $x_n(u, v)$, etc., satisfy conditions (2.1), and

$$E_n = G_n \text{ and } F_n = 0 \text{ almost everywhere.} \tag{6.5}$$

Moreover, we may assume that equations (6.4) carry the points A_1^*, A_2^*, A_3^* into $A_1^{(n)}$, $A_2^{(n)}$, $A_3^{(n)}$ respectively; for if the $A_i^{(n)}$ correspond under (6.4) to B_i^* $(i=1, 2, 3)$ on the unit circle, then by a conformal mapping $u=u(\bar{u}, \bar{v})$, $v=v(\bar{u}, \bar{v})$, we can map the B_i^* on the A_i^*, and by Theorem I the new functions $x(u(\bar{u}, \bar{v}), v(\bar{u}, \bar{v}))$, etc., continue to satisfy conditions (2.1) and (6.5). By (6.3), we have

$$\begin{aligned}\iint_{u^2+v^2<1} \tfrac{1}{2}(E_n + G_n)dudv \\ = \iint_{u^2+v^2<1} (E_nG_n - F_n^2)^{1/2}\, dudv = L(S_n) \leqq L(\Pi_n);\end{aligned} \tag{6.6}$$

hence the first expression in (6.6) is bounded. The sequence $\{S_n\}$ therefore satisfies the hypotheses of Lemma 2, and so there exists a representation $x=x(\theta)$, $y=y(\theta)$, $z=z(\theta)$ of the curve Γ and a subsequence of $\{S_n\}$ (for which we retain the same notation) such that

$$\lim x_n(\theta) = x(\theta), \quad \lim y_n(\theta) = y(\theta), \quad \lim z_n(\theta) = z(\theta)$$

uniformly in θ.

The surface S_n has one representation $x=\bar{x}_n(u, v)$, etc., in which the functions $\bar{x}_n(u, v)$, etc., are monotonic; hence by §1 we know that in the representation (6.4) of S_n the functions $x_n(u, v)$, etc., are monotonic. The sequences $\{x_n(u, v)\}$, $\{y_n(u, v)\}$, $\{z_n(u, v)\}$ are therefore seen to satisfy all the hypotheses of Lemma 1; hence there exist monotonic functions $\bar{x}(u, v)$, $\bar{y}(u, v)$, $\bar{z}(u, v)$ defined on the unit circle such that

$$\begin{aligned}\lim x_n(u, v) = \bar{x}(u, v), \quad \lim y_n(u, v) = \bar{y}(u, v), \\ \lim z_n(u, v) = \bar{z}(u, v)\end{aligned} \tag{6.7}$$

uniformly on the whole circle. Consider now the surface

$$S:\quad x = \bar{x}(u, v),\ y = \bar{y}(u, v),\ z = \bar{z}(u, v),\ u^2 + v^2 \leqq 1;$$

its boundary curve is a limit curve of the Γ_n, since the convergence in (6.7) is uniform on the circumference $u^2+v^2=1$, and since the Γ_n have the unique limit Γ, the boundary curve of S is Γ itself. Moreover, by the semi-continuity of the Lebesgue area we have

$$L(S) \leqq \liminf L(S_n) \leqq \lim L(\Pi_n) = m(\Gamma). \tag{6.8}$$

But since S is bounded by Γ, we have

$$L(S) \geqq a(\Gamma); \tag{6.9}$$

comparing inequalities (6.8), (6.9) and (6.1), we have $L(S)=a(\Gamma)=m(\Gamma)$, which establishes equation (6.3).

By Theorem I, there exists a representation

$$x = x(u, v),\ y = y(u, v),\ z = z(u,v),\ u^2 + v^2 = 1, \tag{6.10}$$

of S for which conditions (2.1) are satisfied, and further

$$E = G \text{ and } F = 0 \text{ almost everywhere.} \tag{6.11}$$

Then†

$$L(S) = \iint_{u^2+v^2<1} (EG - F^2)^{1/2} dudv = \iint_{u^2+v^2<1} \tfrac{1}{2}(E + G) dudv. \tag{6.12}$$

From this it follows that $x(u, v)$, $y(u, v)$, $z(u, v)$ are harmonic. For suppose, e.g., that $x(u, v)$ is not harmonic, and let $\xi(u, v)$ be the harmonic function having the same boundary values as $x(u, v)$. The function ξ minimizes‡ the Dirichlet integral for the given boundary values, and is the unique minimizing function; hence for the surface

$$\bar{S}:\quad x = \xi(u, v),\ y = y(u, v),\ z = z(u, v),\ u^2 + v^2 \leqq 1,$$

we have $\bar{E}+\bar{G}<E+G$. Therefore

$$\begin{aligned} L(\bar{S}) &= \iint_{u^2+v^2<1} (\bar{E}\bar{G} - \bar{F}^2)^{1/2} dudv \leqq \iint_{u^2+v^2<1} \tfrac{1}{2}(\bar{E} + \bar{G}) dudv \\ &< \iint_{u^2+v^2<1} \tfrac{1}{2}(E + G) dudv = L(S) = a(\Gamma). \end{aligned} \tag{6.13}$$

† McShane or Morrey, loc. cit. in §4.

‡ Lebesgue's proof of the minimizing property (Bulletin de la Société Mathématique de France, vol. 41 (1913); p. 48 of the Comptes Rendus) can easily be modified to show that $\xi(u, v)$ minimizes the Dirichlet integral in the class of all functions having the given boundary values and satisfying conditions (2.1).

But S is bounded by Γ, hence $L(S) \geqq a(\Gamma)$, contradicting (6.13). Hence $x(u, v)$, $y(u, v)$, $z(u, v)$ are harmonic.

Since E, F, G are now seen to be continuous, equations (6.11) imply $E=G$, $F=0$ everywhere in the unit circle. By a theorem of Weierstrass we know that if a surface S is so represented that $E=G$, $F=0$, the surface S is minimal if and only if the functions x, y, z are harmonic; these conditions being here satisfied, our surface S is a minimal surface, and the theorem is proved.

Between the present solution of the problem of Plateau and that given by Radó there remains one point of difference. We have not shown that our equations (6.2) carry the circumference of the unit circle in a one-to-one way into the curve Γ. We can however very easily establish this by the same device as was used by Radó,† to whose work we refer the reader.

---

† Radó, loc. cit.; in particular, chapter 2, §3, No. 9.

University of Chicago,
Chicago, Ill.

THE LATIN SQUARE, OR CYCLIC, FUNCTIONS*

BY

E. T. BELL

1. Introduction. Special cases of the Latin square functions defined in this paper have recently come into some prominence in connection with generalizations by Humbert and others (references in §5) of the partial differential equations of mathematical physics. In solving the equations, the functions of $r-1$ independent variables defined by Appell[8]† in 1877 appear, and these in turn are intimately connected with Olivier's[1]† functions $f_0(x), \cdots, f_{r-1}(x)$, whose generating identity is

$$\exp \alpha x = f_0(x) + \alpha f_1(x) + \cdots + \alpha^{r-1} f_{r-1}(x), \tag{1.1}$$

where α is an imaginary rth root of unity,

$$f_j(x) = \sum x^{n_j}/n_j!, \tag{1.2}$$

the summation referring to all integers $n_j \geqq 0$ such that $n_j \equiv j \bmod r$. We shall call r the *base* of $f_j(n)$. Appell's functions A_t can be defined by expanding the left member of the following identity as a power series in α, and reducing the result modulo $\alpha^r - 1$,

$$\exp\left(\sum_{s=1}^{r-1} \alpha^s x_s\right) = \sum_{t=0}^{r-1} \alpha^t A_t(x_1, \cdots, x_{r-1}). \tag{1.3}$$

The r functions $A_t(x_1, \cdots, x_{r-1}) \equiv A_t$ are connected by the identical algebraic relation

$$N(A_0, \cdots, A_{r-1}) = 1, \tag{1.4}$$

where $N(y_0, \cdots, y_{r-1})$ is the norm of the algebraic number

$$y_0 + \alpha y_1 + \cdots + \alpha^{r-1} y_{r-1}.$$

As the partial differential equations mentioned have no immediate physical significance, there is no apparent reason for stopping short of the general case. In a previous paper[7] the functions defined by reducing the left of (3) modulo $P(\alpha)$, where $P(\alpha)$ is any polynomial in α, were introduced and some of their properties discussed. The norm property (1.4) does not hold for these functions, except in the very degenerate case when they become Appell's

* Presented to the Society, March 18, 1933; received by the editors December 27, 1932.

† Numbers refer to bibliography in §5.

It will be interesting to see what replaces the norm property, and how it degenerates in the special case.

We shall see that the generalized norm property is intimately connected with Latin squares. A Latin square of degree n is a square array of n distinct elements such that no element occurs twice in the same column. The number of Latin squares of degree n, no two of which can be derived from one another by a permutation of rows or of columns, will be denoted by $\lambda(n)$. This number has not been determined for general n, and even for small n the labor of a direct determination is prohibitive (see MacMahon[6]). As observed by Cayley,[5] not every Latin square of given degree can be generated by a group of substitutions on the elements of a given row. Thus there exist (even for n small) Latin squares with which no group is associated.

The norm relation is replaced for the generalized functions of r independent variables by $\lambda(r)$ algebraic relations, each of which is derived from a Latin square of degree r. When the functions degenerate to Appell's (based on rth roots of unity), the $\lambda(r)$ relations coalesce in the norm relation, and the single Latin square corresponding to this relation is generated from its first row by the cyclic group of degree r.

Appell's functions are a simple generalization to functions of r independent variables of the circular and hyperbolic functions. The Latin square functions pass at once to the most general situation possible of this kind, namely to the functions of r independent variables constructed from polynomials in the members of sets of r linearly independent solutions of equations of the type

$$\frac{d^r y}{dx^r} + c_1 \frac{d^{r-1} y}{dx^{r-1}} + \cdots + c_r y = 0,$$

where $c_1, \cdots, c_r$ are arbitrary constants, instead of from the degenerate case $c_r = -1, c_j = 0, j \neq r$. The coefficients in the power series for Olivier's functions, on which Appell's are based, are periodic. In the generalized functions periodicity, $\phi(n+r) = \phi(n)$ for all integers n, is replaced by

$$\phi(n + r) + c_1\phi(n + r - 1) + \cdots + c_r\phi(n) = 0,$$

which becomes periodicity in the degenerate case.

All the functions defined are obviously continuous and convergent absolutely for all finite values of the variables.

2. **Generalized Olivier functions.** Consider first the generalization of Olivier's functions. Let

$$P(\alpha) \equiv \alpha^r + c_1\alpha^{r-1} + \cdots + c_r \tag{2.1}$$

be irreducible in the rational domain. Reduction modulo $P(\alpha)$ of the expan-

sion of $\exp \alpha^s x$, where s is an integer, defines the functions $f_j\binom{x}{s}$ uniquely,

$$\exp \alpha^s x = \sum_{j=0}^{r-1} \alpha^j f_j\binom{x}{s}, \tag{2.2}$$

since $P(\alpha)$ is irreducible. We write

$$f_j\binom{x}{1} \equiv f_j(x) \qquad (j = 0, \cdots, r-1). \tag{2.3}$$

The notation in (2.1) is fixed throughout the paper.

The jth fundamental sequence $\phi_j(n)$, $n=0, \pm 1, \pm 2, \cdots$, defined by the difference equation

$$\phi(n+r) + c_1\phi(n+r-1) + \cdots + c_r\phi(n) = 0, \tag{2.4}$$

whose characteristic equation is $P(\alpha)=0$, is determined by

$$\phi_j(k) = \delta_j^k \text{ (Kronecker delta)}, \quad j, k = 0, \cdots, r-1. \tag{2.5}$$

The $\phi_j(n)$ are a set of r linearly independent solutions of (2.4), and the general solution $\phi(n)$ is

$$\phi(n) = \sum_{j=0}^{r-1} \phi(j)\phi_j(n). \tag{2.6}$$

The notations in (2.4), (2.5) are fixed henceforth.

For all integers n we have

$$\alpha^n = \sum_{j=0}^{r-1} \alpha^j \phi_j(n). \tag{2.7}$$

Hence, by (2.2),

$$f_j\binom{x}{s} = \sum_{n=0}^{\infty} \phi_j(sn)\frac{x^n}{n!} \qquad (j = 0, \cdots, r-1). \tag{2.8}$$

To find the differential equation satisfied by the functions (2.8), let $P_s(\alpha)$ be the polynomial with leading coefficient unity whose roots are the sth powers of the roots of $P_1(\alpha) \equiv P(\alpha)$,

$$P_s(\alpha) \equiv \beta^r + c_1(s)\beta^{r-1} + \cdots + c_r(s) \qquad (\beta \equiv \alpha^s). \tag{2.9}$$

Then, by (2.2), (2.9), r linearly independent solutions of

$$\frac{d^r y}{dx^r} + c_1(s)\frac{d^{r-1}y}{dx^{r-1}} + \cdots + c_r(s)y = 0 \tag{2.10}$$

are the functions (2.8), and the general solution of (2.10) is

$$y = \sum_{j=0}^{r-1} k_j f_j \binom{x}{s}, \tag{2.11}$$

where the k_j are arbitrary constants.

The exponential forms of the functions (2.8), corresponding to those of the circular, hyperbolic, and Olivier functions, are obtained at once from (2.2). If $\alpha_0, \cdots, \alpha_{r-1}$ are the roots of $P(\alpha)=0$, and α^{ij} denotes the cofactor of α_i^j in the determinant

$$D(\alpha) \equiv \begin{vmatrix} 1 & \alpha_0 & \cdots & \alpha_0^{r-1} \\ 1 & \alpha_1 & \cdots & \alpha_1^{r-1} \\ \cdot & \cdot & \cdot & \cdot \\ 1 & \alpha_{r-1} & \cdots & \alpha_{r-1}^{r-1} \end{vmatrix},$$

we have

$$D(\alpha) f_j \binom{x}{s} = \sum_{t=0}^{r-1} \alpha^{tj} \exp \alpha_t x^s, \tag{2.12}$$

since $D(\alpha) \neq 0$, $P(\alpha)$ being irreducible.

Corresponding to the period recurrence of the derivatives of the circular, hyperbolic, and Olivier functions, we have here

$$\frac{d^t}{dx^t} f_k \binom{x}{s} = \sum_{j=0}^{r-1} \phi_k(st+j) f_j \binom{x}{s} \quad (k = 0, \cdots, r-1), \tag{2.13}$$

on differentiating (2.2) t times and applying (2.7).

Applying (2.7) to the product of $\exp \alpha^s x$ and $\exp \alpha^s y$, we get the addition theorems

$$f_j \binom{x+y}{s} = \sum_{p=0}^{r-1} \sum_{k=0}^{r-1} \phi_j(p+k) f_p \binom{x}{s} f_k \binom{y}{s}. \tag{2.14}$$

There is no algebraic addition theorem with respect to s.

Let α be any root of $P(\alpha)=0$. Then

$$\exp [x\alpha^s P(\alpha)] = 1,$$

and hence, by (22), the identical algebraic relations between the functions are obtained by reducing the expression on the left of the following, modulo $P(\alpha)$, to that on the right ($c_0 = 1$),

$$\prod_{k=0}^{r-1} \left[\sum_{j_k=0}^{r-1} \alpha^{j_k} f_{j_k} \binom{c_k x}{s+r-k} \right] = \sum_{j=0}^{r-1} \alpha^j N_j \binom{x}{s}; \tag{2.15}$$

the relations are

$$N_0\binom{x}{s} = 1, \quad N_p\binom{x}{s} = 0 \qquad (p = 1, \cdots, r-1). \tag{2.16}$$

For Olivier's functions it is easily seen that (2.16) are equivalent to the single norm relation (the last $r-1$ relations are absent).

If $P(\alpha)$ is such that, for some integer $s>0$, $c_1(s)=0$ in (2.9), the functions $f_j\binom{x}{s}$, $j=0, \cdots, r-1$, are more simply connected. Let the roots of $P_s(\alpha)=0$ be $\beta_0, \cdots, \beta_{r-1}$. If β is any one of the roots, we may define functions $g_j\binom{x}{t}$ by the process for (2.2) with $P(\alpha)$ replaced by the right of (2.9),

$$\exp \beta^t x = \sum_{j=0}^{r-1} \beta^j g_j\binom{x}{t}. \tag{2.17}$$

Apply (2.7) to $\beta^j \equiv \alpha^{js}$. Then

$$f_k\binom{x}{st} = \sum_{j=0}^{r-1} \phi_k(js) g_j\binom{x}{t} \qquad (k = 0, \cdots, r-1); \tag{2.18}$$

$$g_j\binom{x}{1} \equiv g_j(x).$$

If now $c_1(s)=0$, $\beta_0 + \cdots + \beta_{r-1}=0$, and $\exp(\beta_0+\cdots+\beta_{r-1})x=1$. Hence

$$M_0(g_0(x), \cdots, g_{r-1}(x)) = 1, \tag{2.19}$$

where $M_0(y_0, \cdots, y_{r-1})$ is the norm of $y_0+y_1\beta+\cdots+y_{r-1}\beta^{r-1}$. If further $s=1$, (2.16) hence becomes

$$N_0(f_0(x), \cdots, f_{r-1}(x)) = 1, \tag{2.20}$$

where $N_0(y_0, \cdots, y_{r-1})$ is the norm of $y_0+y_1\alpha+\cdots+y_{r-1}\alpha^{r-1}$. A linear transformation on the $f_i(x)$ will always reduce (2.16) when $s=1$ to (2.20).

Since there are precisely r functions $f_j\binom{x}{s}$ of the single variable x, they must be connected by $r-1$ relations. These are contained in the r relations (2.12), which are not independent, or in the equivalent dependent set obtained from (2.2) by putting $\alpha_0, \cdots, \alpha_{r-1}$ successively for α. The dependence for the last set of r is evident from $-c_1(s)=\alpha_0^s+\cdots+\alpha_{r-1}^s$; $c_1(s)$ is a rational function of $c_1, \cdots, c_{r-1}$. The $r-1$ independent relations are transcendental.

3. Functions with periodic coefficients. A special case of the functions (2.3) is of particular interest as it can be completely specified with remarkable simplicity. In a previous note[10] it was shown that the only difference equations (2.4) whose solutions have the proper additive period m (integer >0) are those in which $r=\tau(m)$, the totient (Euler's function) of m, and $P(\alpha)=0$ is

the equation whose roots are the $\tau(m)$ primitive mth roots of unity. In this section m is a constant integer >0, $r=\tau(m)$, and

$$\alpha^r + c_1\alpha^{r-1} + \cdots + c_r = 0 \tag{3.1}$$

is the equation for the primitive mth roots of unity. All of §2 necessarily holds in this case, with special features not valid in §2. The notation is as before; in particular the general solution of

$$\phi(n + r) + c_1\phi(n + r - 1) + \cdots + c_r\phi(n) = 0 \tag{3.2}$$

is $\phi(n)$. The sequence $\phi(n)$ $(n=0, \pm 1, \cdots)$ is determined by (3.2) when $\phi(0), \cdots, \phi(r-1)$ are given constants.

From what has just been recalled it follows that the only functions

$$f(x) \equiv \sum_{n=0}^{\infty}\psi(n)x^n/n!, \tag{3.3}$$

in which $\psi(n)$ has the proper additive period m and is determined by a linear difference equation with constant coefficients, are those in which $\psi=\phi$,

$$f(x) = \sum_{n=0}^{\infty}\phi(n)x^n/n!,$$

and hence

$$f(x) = \sum_{t=0}^{m-1}\phi(t)\left[\sum_{n=0}^{\infty}\frac{x^{nm+t}}{(nm + t)!}\right]. \tag{3.4}$$

The functions in square brackets, say

$$h_t(x) \equiv \sum_{n=0}^{\infty}\frac{x^{nm+t}}{(nm + t)!}, \tag{3.5}$$

are the m Olivier functions to the base m; see (1.2). Hence, by (2.6), the general function (3.3) with periodic coefficients of the kind described is

$$f(x) = \sum_{j=0}^{r-1}\phi(j)\left[\sum_{t=0}^{m-1}\phi_j(t)h_t(x)\right]. \tag{3.6}$$

Consider the functions in the square brackets in (3.6),

$$H_j(x) \equiv \sum_{t=0}^{m-1}\phi_j(t)h_t(x) \qquad (j = 0, \cdots, r - 1). \tag{3.7}$$

The generating identity is

$$\exp \alpha x = \sum_{j=0}^{r-1}\alpha^j H_j(x),$$

where α is any root of (3.1). Hence, taking the nth derivative, we get

$$\frac{d^n}{dx^n}H_j(x) = \sum_{i=0}^{r-1}\phi_j(n+i)H_i(x); \tag{3.8}$$

and therefore, by the periodicity of ϕ_j,

$$\frac{d^{km+n}}{dx^{km+n}}H_j(x) = \frac{d^n}{dx^n}H_j(x) \tag{3.9}$$

$$(n = 0, \cdots, m-1; j = 0, \cdots, r-1)$$

for all integers $k \geqq 0$. Thus the derivatives of the $H_j(x)$ recur with the period m. Since $\phi(0), \cdots, \phi(r-1)$ in (3.6) are arbitrary constants, (3.9) implies

$$\frac{d^{km+n}}{dx^{km+n}}f(x) = \frac{d^n}{dx^n}f(x), \tag{3.10}$$

and $f(x)$ is the most general function with recurring derivatives of period m.

Consider next the functions (3.3) in which $\psi(n)$ has the proper multiplicative period $m+1$ (m integer >0), and in which $\psi(n)$ is determined by a linear difference equation with constant coefficients. It follows from the theorem recalled for additive periodicity that the only such $\psi(n)$ with multiplicative period $m+1>1$ are the $\phi(n)$ defined by (3.1) as before. Hence the properties of these functions follow from those just discussed.

4. Latin square functions. In this section the notation is as in §2. We shall need particularly (2.1)–(2.4). As a basis for the numbers of the field $K(\alpha)$ we shall take $1, \alpha, \cdots, \alpha^{r-1}$, and we shall denote the element of $K(\alpha)$ whose coordinates are $x_0, \cdots, x_{r-1}$ by (x),

$$(x) \equiv (x_0, \cdots, x_{r-1}) \equiv x_0 + \alpha x_1 + \cdots + \alpha^{r-1}x_{r-1}. \tag{4.1}$$

The sum of (x), (y) may be written either as $(x)+(y)$ or $(x+y)$,

$$(x+y) = (x_0 + y_0, \cdots, x_{r-1} + y_{r-1}); \tag{4.2}$$

their product, $(x)(y)$ or (xy), is

$$\begin{aligned}(xy) &= ((xy)_0, \cdots, (xy)_{r-1}),\\ (xy)_j &\equiv \sum_{i=0}^{r-1}\sum_{p=0}^{r-1}\phi_j(i+p)x_iy_p.\end{aligned} \tag{4.3}$$

More generally, the product of any finite number of elements $(x), (y), \cdots,$ (z) of $K(\alpha)$, in a similar notation, is given by

$$\begin{gathered}(xy\cdots z)_j \equiv \sum\phi_j(i+p+\cdots+t)x_iy_p\cdots z_t,\\ 0 \leqq i, p, \cdots, t \leqq r-1.\end{gathered} \tag{4.4}$$

The element $(x)'$ of $K(\alpha)$ defined by

$$
(4.5)\qquad
\begin{aligned}
(x)' &\equiv (x) - x_0 + x_r\alpha^r \equiv (x_0', \cdots, x_{r-1}'),\\
x_0' &= -c_r x_r,\quad x_k' = x_k - c_{r-k}x_r \qquad (k = 1, \cdots, r-1),\\
(x)' &= \alpha x_1 + \alpha^2 x_2 + \cdots + \alpha^r x_r
\end{aligned}
$$

will be called the *curtate* of (x). Accents as in $(x)'$, $(y)'$, $\cdots$ shall denote the curtates of the corresponding (x), (y), $\cdots$.

The *Latin square functions of degree r* in the independent variables $x_1, \cdots, x_r$ are denoted by $L_j(x_1, \cdots, x_r)$, and are defined by the identity (4.6), in which the right is the reduction modulo $P(\alpha)$ of the expansion of the left as a power series in α,

$$
(4.6)\qquad \exp (x)' \equiv \sum_{j=0}^{r-1} \alpha^j L_j(x_1, \cdots, x_r).
$$

To find the algebraic relations between the L_j mentioned in §1, we proceed as described presently from the Latin square (4.7) to its "bordered mate" (4.8). We assume $r \geqq 1$. Let $x_1^{(i)}, \cdots, x_r^{(i)}$ be the ith row of the Latin square (4.7) of degree r constructed from $x_1, \cdots, x_r$, so that $x_1^{(i)}, \cdots, x_r^{(i)}$ is some permutation of $x_1, \cdots, x_r$:

$$
(4.7)\qquad
\begin{array}{cccc}
x_1^{(1)}, & x_2^{(1)}, & \cdots, & x_r^{(1)}\\
x_1^{(2)}, & x_2^{(2)}, & \cdots, & x_r^{(2)}\\
\cdot & \cdot & \cdot & \cdot\\
x_1^{(r)}, & x_2^{(r)}, & \cdots, & x_r^{(r)}.
\end{array}
$$

Write $-s \equiv x_1 + x_2 + \cdots + x_r$. Multiply the elements in the jth column, $j > 1$, of (4.7) by c_{j-1}. Apply $\alpha^r, \alpha^{r-1}, \cdots, \alpha$ as top border to the result, and $s, c_1 s, c_2 s, \cdots, c_{r-1}s$ as a bottom border:

$$
(4.8)\qquad
\begin{array}{cccc}
\alpha^r, & \alpha^{r-1}, & \cdots, & \alpha\\
\hline
x_1^{(1)}, & c_1 x_2^{(1)}, & \cdots, & c_{r-1}x_r^{(1)}\\
x_1^{(2)}, & c_1 x_2^{(2)}, & \cdots, & c_{r-1}x_r^{(2)}\\
\cdot & \cdot & \cdot & \cdot\\
x_1^{(r)}, & c_1 x_2^{(r)}, & \cdots, & c_{r-1}x_r^{(r)}\\
\hline
s, & c_1 s, & \cdots, & c_{r-1}s
\end{array}
$$

Consider the rows of (4.8) as vectors and take the inner product of the vector whose coordinates are the top border by each of the remaining $r+1$. The sum of these $r+1$ inner products vanishes, as it is $(x_1 + \cdots + x_r)\, P(\alpha)$, from the

construction of (4.8). These products are the curtates

$$(4.9)\qquad \begin{aligned}&(c_{r-1}x_r^{(i)}, c_{r-2}x_{r-1}^{(i)}, \cdots, c_1x_2^{(i)}, x_1^{(i)})' \qquad (i = 1, \cdots, r),\\ &(c_{r-1}s, c_{r-2}s, \cdots, c_1s, s)'.\end{aligned}$$

To simplify the writing, let the $r+1$ curtates in (4.9) be equal respectively to

$$(4.10)\qquad \begin{aligned}&(y_1^{(i)}, y_2^{(i)}, \cdots, y_r^{(i)})' \qquad (i = 1, \cdots, r),\\ &(y_1^{(r+1)}, y_2^{(r+1)}, \cdots, y_r^{(r+1)})'.\end{aligned}$$

Then

$$\exp\left[\sum_{p=1}^{r+1}(y_1^{(p)}, \cdots, y_r^{(p)})'\right] = 1$$

and hence, by (4.6),

$$(4.11)\qquad \prod_{p=1}^{r+1}\left[\sum_{i_p=0}^{r-1}\alpha^{i_p}L_{i_p}(y_1^{(p)}, \cdots, y_r^{(p)})\right] = 1.$$

When distributed and reduced modulo $P(\alpha)$, the left of (4.11) is of the form $N_0+\alpha N_1+\cdots+\alpha^{r-1}N_{r-1}$, where N_j is a homogeneous polynomial of degree $r+1$ in functions $L_0, L_1, \cdots, L_{r-1}$ whose variables are given in (4.10). For the moment the structure of the N_j need not be considered. Starting then with the particular Latin square (4.7), we reach the identical relations

$$(4.12)\qquad N_0 = 1,\ N_j = 0 \qquad (j = 1, \cdots, r-1).$$

We indicate the structure of the N's presently.

From (4.5), (4.6) we find explicit forms for the L_j. The expression for the L_j corresponding to (2.12) is obvious and can be omitted. Let $\theta_0, \cdots, \theta_{r-1}$ be the r conjugates of $(x)'$, including $(x)'$. Form the equation

$$(4.121)\qquad \theta^r + b_1\theta^{r-1} + \cdots + b_r = 0 \qquad (\theta = \theta_0, \cdots, \theta_{r-1})$$

whose roots are these conjugates. Then $b_j \equiv b_j(x_1, \cdots, x_r)$ is a homogeneous polynomial of degree j in $x_1, \cdots, x_r$, whose coefficients are polynomials in $c_1, \cdots, c_r$ with rational integer coefficients. Similarly to the discussion for (2.3)–(2.5) we consider the difference equation

$$(4.13)\qquad \xi(n + r) + b_1\xi(n + r - 1) + \cdots + b_r\xi(n) = 0,$$

whose characteristic equation is (4.121). The r fundamental sequences $\xi_j(n)$ for (4.13) are determined by

$$(4.14)\qquad \xi_j(k) = \delta_j^k \qquad (j, k = 0, \cdots, r-1);$$

the general solution $\xi(n)$ is

$$\xi(n) = \sum_{j=0}^{r-1}\xi(j)\xi_j(n); \tag{4.15}$$

and we have

$$\theta^n = \sum_{j=0}^{r-1}\theta^j\xi_j(n). \tag{4.16}$$

The powers of θ on the right of (4.16) must be reduced modulo $P(\alpha)$ independently. Let

$$\theta^j = p_{j0} + \alpha p_{j1} + \cdots + \alpha^{r-1}p_{j,r-1} \quad (j = 1, \cdots, r-1). \tag{4.17}$$

Then $p_{ji} \equiv p_{ji}(x_1, \cdots, x_r)$ is a polynomial in $x_1, \cdots, x_r$ whose coefficients are polynomials in $c_1, \cdots, c_r$ with rational integer coefficients. From (4.16) we now have

$$\theta^n = \sum_{j=0}^{r-1}\alpha^j\left[\sum_{i=0}^{r-1}p_{ij}\xi_i(n)\right]; \tag{4.18}$$

hence, by (4.6),

$$L_j(x_1, \cdots, x_r) = \sum_{i=0}^{r-1}p_{ij}\sum_{n=0}^{\infty}\frac{\xi_i(n)}{n!} \quad (j = 0, \cdots, r-1). \tag{4.19}$$

The p_{ij} are defined by (4.17), and the $\xi_j(n)$ are the fundamental solutions of (4.13).

Since the variables $x_1, \cdots, x_r$ are independent, the differential relations of §2 go over, by (4.6), to corresponding relations for the L_j. Thus from (2.10), (4.6) we have

$$\left[\frac{\partial^r}{\partial x_s{}^r} + c_1(s)\frac{\partial^{r-1}}{\partial x_s{}^{r-1}} + \cdots + c_r(s)\right]L_j(x_1, \cdots, x_r) = 0 \tag{4.20}$$

$$(j = 0, \cdots, r-1;\ s = 1, \cdots, r);$$

and corresponding to (2.13),

$$\frac{\partial^t}{\partial x_s{}^t}L_k(x_1, \cdots, x_r) = \sum_{j=0}^{r-1}\phi_k(st+j)L_j(x_1, \cdots, x_r); \tag{4.21}$$

whence

$$\left(\frac{\partial^m}{\partial x_n^m} - \frac{\partial^n}{\partial x_m^n}\right)L_k(x_1, \cdots, x_r) = 0 \quad (m, n = 1, \cdots, r).$$

The expressions for the L_k as polynomials in the functions defined in (2.8) follow at once from (4.4)–(4.6),

$$L_k(x_1, \cdots, x_r) = \sum \phi_k(j_0 + \cdots + j_{r-1}) f_{j_0}\binom{x_0'}{0} \cdots f_{j_{r-1}}\binom{x_{r-1}'}{r-1},$$

the sum extending to all $0 \leqq j_0, \cdots, j_{r-1} \leqq r-1$. The N's in (4.12) have a similar structure in terms of L's. The addition theorems are of the same type, but simpler,

$$F_t(x_1 + y_1, \cdots, x_r + y_r) = \sum \phi_t(j + k) F_j(x_1, \cdots, x_r) F_k(y_1, \cdots, y_r),$$

summed for $0 \leqq j, k \leqq r-1$.

From this point on, the connection with partial differential equations is of the same kind as that for the Appell functions and the equations discussed by Humbert and others in the papers cited in §5. The note [14] sufficiently indicates the start.

5. **References.** Several of the following papers contain further references to the literature of Appell's functions and their connection with differential equations. The references are given in chronological order. Humbert (loc. cit., p. 153) attributes Olivier's functions to Yvon Villarceau, without stating the reference.

1. L. Olivier, *Bemerkungen über eine Art von Funktionen* . . . , Crelle's Journal, vol. 2 (1827), pp. 243–251.

2. J. W. L. Glaisher, *On functions with recurring derivatives*, Proceedings of the London Mathematical Society, vol. 4 (1872), pp. 113–116.

3. P. Appell, *Sur certaines fonctions analogues aux fonctions circulaires*, Comptes Rendus, Paris, vol. 84 (1877), pp. 1378–1380.

4. J. W. L. Glaisher, *Functions analogous to the sine and cosine*, Quarterly Journal, vol. 16 (1879), pp. 15–33.

5. A. Cayley, *On Latin squares*, Messenger of Mathematics, vol. 19 (1890), pp. 135–137 (Collected Papers, vol. 13, No. 903).

6. P. A. MacMahon, *Combinatorial Analysis*, vol. 1, 1915.

7. E. T. Bell, *Periodic functions of n variables connected with an algebraic number field of degree n*, Quarterly Journal, vol. 50 (1927), pp. 314–328.

8. P. Humbert, *Sur une généralisation de l'équation de Laplace*, Journal des Mathématiques, (9), vol. 8 (1929), pp. 145–159.

9. D. V. Jonescu, *Sur une équation aux dérivées partielles du troisième ordre*, Bulletin de la Société Mathématique de France, vol. 58 (1930), pp. 224–229.

10. E. T. Bell, *Periodic recurring series*, Proceedings of the National Academy of Sciences, vol. 16 (1930), pp. 750–752.

11. J. Devisme, Comptes Rendus, Paris, vol. 193 (1931), pp. 981–983, 825–828; ibid., vol. 194, pp. 516–519.

12. P. Humbert, *On Appell's function $P(\theta, \phi)$*, Proceedings of the Edinburg Mathematical Society, (2), vol. 3(1932), pp. 53–55.

13. J. Devisme, *Sur la fonction génératrice de la fonction $P(m\theta, n\phi)$ d'Appell*, Académie Royale de Belgique, Bulletin, (5), vol. 18 (1932), pp. 505–506.

14. E. T. Bell, *A Laplacian equation*, American Mathematical Monthly, vol. 39 (1932), pp. 515–517.

California Institute of Technology,
Pasadena, Calif.

SUFFICIENT CONDITIONS IN THE PROBLEM OF THE CALCULUS OF VARIATIONS IN n-SPACE IN PARAMETRIC FORM AND UNDER GENERAL END CONDITIONS*

BY

SUMNER BYRON MYERS

1. **Introduction.** Sufficient conditions in the general problem of the calculus of variations .in parametric form are given here. The results are in terms of the characteristic roots of a linear boundary value problem, and are in close relation to the conditions recently given by Morse† in the corresponding problem in non-parametric form.

An important feature of the results is that the usual "non-tangency" hypothesis is not made. For example, if these results were applied to the problem of minimizing an integral along curves joining a point to a manifold, we would obtain sufficient conditions for a minimum even in the case that the minimizing curve is tangent to the manifold.

The essential idea in the methods used in the paper is the treatment of the parametric problem as the limiting case of a series of non-singular non-parametric problems by means of a suitable modification of the integrand. Although they lack the geometric invariance of methods now being developed by Morse, in which the parametric problem is approximated by means of a series of parametric problems of the same nature as the original problem, the methods and results of this paper derive advantage from the non-singularity of the approximating non-parametric problems and from the fact that the cases of "non-tangency" and "tangency" are treated together. The work of the author and that of Morse are thus complementary, and constitute the first complete treatment of sufficient conditions in the general parametric problem.

---

* Presented to the Society, March 26, 1932; received by the editors April 19, 1932.

† Certain results in the following papers will be used.

Morse, *Sufficient conditions in the problem of Lagrange with variable end conditions*, American Journal of Mathematics, vol. 53 (1931), pp. 517–546.

Morse and Myers, *The problems of Lagrange and Mayer with variable end points*, Proceedings of the American Academy of Arts and Sciences, vol. 66 (1931), pp. 235–253.

Bliss, *Jacobi's condition for problems of the calculus of variations in parametric form*, these Transactions, vol. 17 (1916), pp. 195–206.

Further references can be found in the three papers just cited.

2. The Euler equations and the transversality conditions. In the space of the variables

$$(x) = (x_1, \cdots, x_n)$$

let there be given an ordinary arc g

$$(2.1) \qquad x_i = \bar{x}_i(t), \quad t^{(1)} \leqq t \leqq t^{(2)} \qquad (i = 1, \cdots, n),$$

of class C'.

We consider ordinary arcs of class D' neighboring g. The initial and final end points of such arcs will be denoted respectively by

$$(x^s) = (x_1^s, \cdots, x_n^s) \qquad (s = 1, 2)$$

and the end values of the parameter t will be denoted respectively by t^s ($s=1, 2$), where $s=1$ at the initial end point and $s=2$ at the final end point. An ordinary arc of class D' neighboring g will be said to be *admissible* if its end points are given for some value of

$$(\alpha) = (\alpha_1, \cdots, \alpha_r)$$

by the functions

$$(2.2) \qquad x_i^s = x_i^s(\alpha_1, \cdots, \alpha_r), \quad 0^* \leqq r \leqq 2n \quad (i = 1, \cdots, n; s = 1, 2).$$

These functions of (α) are of class C'' for (α) near (0) and reduce to the end points of g for $(\alpha)=(0)$. We assume that the functional matrix of the functions in (2.2)

$$\|x_{ih}^s\| \quad (h = 1, \cdots, r; i = 1, \cdots, n; s = 1, 2)$$

is of rank r for $(\alpha)=(0)$. Here and henceforth the subscript h attached to x_i^s shall denote differentiation with respect to α_h.

We seek conditions under which the arc g and the set $(\alpha)=(0)$ afford a minimum to the expression

$$(2.3) \qquad J = \int_{t^1}^{t^2} F(x, \dot{x})dt + \theta(\alpha)$$

among sets (α) near (0) and admissible arcs neighboring g with end points determined by these sets (α). The function $F(x, \dot{x})$ is defined for (x) in an open region containing g and for $(\dot{x})$ any set not (0), and is to be of class C'''. The function θ is to be of class C'' for (α) near (0).

Furthermore, the function F is to satisfy the usual homogeneity relation

$$(2.4) \qquad F(x, k\dot{x}) \equiv kF(x, \dot{x}), \qquad k > 0.$$

* The case $r=0$ yields the fixed end point problem. This case will be treated separately at the end of the paper, so that until then we shall assume that $r>0$.

Certain necessary conditions are obtained immediately by treating the problem as a non-parametric problem of minimizing J among curves of class D' in the $(n+1)$-space of the variables (t, x) whose end points satisfy (2.2) and the conditions $t = t^{(s)}$.*

THEOREM 1. *If g affords a minimum to J in the problem, then along g the following equations must be satisfied*:

$$\frac{d}{dt}\left[\frac{\partial F}{\partial \dot{x}_i}\right] - \frac{\partial F}{\partial x_i} = 0 \qquad (i = 1, \cdots, n), \tag{2.5}$$

while the following transversality relations must hold:

$$\left[\frac{\partial F}{\partial \dot{x}_i}\overset{s}{x}_{ih}\right]_1^2 + \theta_h = 0\dagger \quad (h = 1, \cdots, r;\ i = 1, \cdots, n). \tag{2.6}$$

We shall now state and prove a theorem which will be useful later.

THEOREM 2. *For an arbitrary set of functions $\eta_i(t)$ of class D' such that $\eta_i(t^{(s)}) = x^s_{ih}u_h$ $(i = 1, \cdots, n;\ h = 1, \cdots, r;\ s = 1, 2)$ for some set of numbers $(u_1, \cdots, u_r)$, there exists a one-parameter family of admissible arcs*

$$x_i = x_i(t, e), \quad \alpha_h = \alpha_h(e) \quad (i = 1, \cdots, n;\ h = 1, \cdots, r), \tag{2.7}$$

containing g for $e = 0$, with $\eta_i(t)$ and u_h as its respective variations; *that is, the functions in* (2.7) *will have the following properties*:

$$\begin{aligned} &x_i(t, 0) = \bar{x}_i(t), \\ &x_i(\overset{s}{t}, e) = \overset{s}{x}_i[\alpha(e)], \\ &\alpha_h(0) = 0, \quad x_{ie}(t, 0) = \eta_i(t), \quad \alpha_h'(0) = u_h \\ &\qquad\qquad (i = 1, \cdots, n;\ h = 1, \cdots, r;\ s = 1, 2). \end{aligned} \tag{2.8}$$

Furthermore, the functions $x_i(t, e)$ and $x_{ie}(t, e)$ are continuous and have continuous derivatives with respect to e for e near 0 *and t in the interval $t^{(1)} \leqq t \leqq t^{(2)}$, while the functions $x_{it}(t, e)$ and $x_{iet}(t, e)$ have the same properties except possibly at the values of t defining the corners of (η). The functions $\alpha_h(e)$ are of class C''.*

For the following is such a family:

$$\begin{aligned} x_i = {}& \bar{x}_i(t) + e[\eta_i(t) - \eta_i^1 h^2(t) - \eta_i^2 h^1(t)] + [x_i^1(eu) - \bar{x}_i^1]h^2(t) \\ & + [x_i^2(eu) - \bar{x}_i^2]h^1(t), \\ \alpha_h = {}& eu_h \qquad\qquad (h = 1, \cdots, r;\ i = 1, \cdots, n), \end{aligned} \tag{2.9}$$

* See Morse and Myers, p. 245, loc. cit.

† Here and henceforth $[\ \ ]_1^2$ shall mean the difference between the value of the bracket evaluated for $s = 2$ and $(x, \dot{x})$ at the final end point of g, and the corresponding evaluation at the initial end point of g. Also, an index repeated in the same term shall always mean summation with respect to that index. The notation θ_h stands for $(\partial\theta/\partial\alpha_h)(0)$.

where $h^1(t)$, $h^2(t)$ are any functions of class C' such that

$$h^1(t^{(1)}) = 0, \quad h^1(t^{(2)}) = 0, \quad h^2(t^{(1)}) = 0, \quad h^2(t^{(2)}) = 0,$$

while η_i^s is an abbreviation for $\eta_i(t^{(s)})$ and $\dot{x}_i^s$ is an abbreviation for $\dot{x}_i(t^{(s)})$.

3. **The accessory boundary problem and a further necessary condition.** We assume now that g is an extremal satisfying the transversality conditions (2.6). We shall use permanently the notations

$$\begin{aligned} \eta_i(t) &= x_{i\epsilon}(t, 0), \\ \overset{s}{\eta_i} &= \eta_i(t^{(s)}), \\ u_h &= \alpha_h'(0) \qquad (i = 1, \cdots, n;\ s = 1, 2;\ h = 1, \cdots, r). \end{aligned}$$

Consider now a family of admissible arcs of form (2.7) satisfying the first three conditions of (2.8) and possessing the differentiability properties of Theorem 2. If we consider this family momentarily as a family of arcs in (t, x)-space satisfying the end conditions

$$\begin{aligned} \overset{s}{x_i} &= \overset{s}{x_i}(\alpha), \\ t^s &= t^{(s)} \qquad (i = 1, \cdots, n;\ s = 1, 2), \end{aligned} \tag{3.1}$$

then we can apply known results* to obtain the second variation of J along this family. We find that

$$J''(0) = b_{hk}u_hu_k + 2\int_{t^{(1)}}^{t^{(2)}} \omega(\eta, \dot{\eta})dt \qquad (h, k = 1, \cdots, r), \tag{3.2}$$

where

$$2\omega(\eta, \dot{\eta}) = \frac{\partial^2 F}{\partial \dot{x}_i \partial \dot{x}_j}\dot{\eta}_i\dot{\eta}_j + 2\frac{\partial^2 F}{\partial \dot{x}_i \partial x_j}\dot{\eta}_i\eta_j + \frac{\partial^2 F}{\partial x_i \partial x_j}\eta_i\eta_j \ (i, j = 1, \cdots, n) \tag{3.3}$$

and

$$b_{hk} = \left[\frac{\partial F}{\partial \dot{x}_i}\overset{s}{x}_{ihk}\right]_1^2 + \theta_{hk} \ \ (h, k = 1, \cdots, r;\ i = 1, \cdots, n). \tag{3.4}$$

With the idea of dominating the sign of the second variation by adding new terms, we are led to consider the accessory problem of minimizing

$$I(\eta, u, \sigma) = b_{hk}u_hu_k + \int_{t^{(1)}}^{t^{(2)}} [2\omega - \sigma(\dot{\eta}_i\dot{\eta}_i + \eta_i\eta_i)]dt \qquad (i = 1, \cdots, n;\ h, k = 1, \cdots, r) \tag{3.5}$$

for a given number σ, relative to constants (u) and functions (η) of class D' satisfying

* See Morse, p. 521, loc. cit.

$$\overset{s}{\eta}_i = \overset{s}{x}_{ih}u_h \quad (i = 1, \cdots, n;\ s = 1, 2;\ h = 1, \cdots, r). \tag{3.6}$$

A *solution* (η), (u) of this new minimum problem in which the functions (η) are of class C'' must satisfy the conditions of the following boundary value problem:

$$\frac{d}{dt}\left[\frac{\partial\Omega}{\partial\dot\eta_i}\right] - \frac{\partial\Omega}{\partial\eta_i} = 0 \qquad (i = 1, \cdots, n), \tag{3.7}$$

$$b_{hk}u_k + \left[\frac{\partial\Omega}{\partial\dot\eta_i}\overset{s}{x}_{ih}\right]_1^2 = 0 \qquad (h, k = 1, \cdots, r), \tag{3.8}$$

$$\overset{s}{\eta}_i = \overset{s}{x}_{ih}u_h \qquad (s = 1, 2), \tag{3.9}$$

where

$$2\Omega(\eta, \dot\eta, \sigma) = 2\omega(\eta, \dot\eta) - \sigma(\dot\eta_i\dot\eta_i + \eta_i\eta_i) \quad (i = 1, \cdots, n). \tag{3.10}$$

This boundary problem we shall call the *accessory boundary problem*. By a *solution* of the accessory boundary problem is meant a set of functions $\eta_i(t)$ of class C'' which with constants (u) and σ satisfy the conditions of the problem. A *characteristic* solution is one for which $(\eta) \not\equiv (0)$.

The corresponding value of σ will be called a *characteristic root*.

The following lemma and theorem can be proved in a manner similar to that used by Morse in his proof of the corresponding results for the non-parametric problem.† In the proof of Theorem 3, Theorem 2 must be used.

LEMMA 1. *If (η) is a characteristic solution with constants (u) and σ, $I(\eta, u, \sigma) = 0$.*

THEOREM 3. *If g furnishes a minimum for the given problem, there can exist no characteristic root $\sigma < 0$.*

4. **The function $J(\alpha)$ and the quadratic form $H(u, \sigma)$.** By the *Legendre sufficient condition* we mean the condition

$$\frac{\partial^2 F}{\partial\dot x_i\partial\dot x_j}\pi_i\pi_j > 0 \qquad (i, j = 1, \cdots, n) \tag{4.1}$$

along g, for all sets $(\pi) \neq (0)$ and not proportional to $(\partial\dot x/\partial t)$.

By the *Weierstrass sufficient condition* we mean the condition

$$E(x, \dot x, \dot y) \equiv F(x, \dot y) - \dot y_i\frac{\partial F}{\partial\dot x_i}(x, \dot x) > 0 \quad (i = 1, \cdots, n) \tag{4.2}$$

for all (x), $(\dot x)$ on g, and for all $(\dot y) \neq (0)$ and not proportional to $(\bar x')$.

† See Morse, p. 254, loc. cit.

We shall assume henceforth that g is an extremal along which the Legendre sufficient condition holds. Among the well known consequences of this assumption are the following:

(1) The determinant

$$\begin{vmatrix} \dfrac{\partial^2 F}{\partial \dot{x}_i \partial \dot{x}_j} & \bar{x}_i' \\ \bar{x}_j' & 0 \end{vmatrix} \neq 0$$

along g.

(2) The functions $\bar{x}_i(t)$ are of class C'''.

(3) The characteristic determinant

$$\left| \frac{\partial^2 F}{\partial \dot{x}_i \partial \dot{x}_j} - \sigma \delta_i^j \right|$$

does not vanish for $\sigma < 0$, where δ_i^j is the Kronecker delta.

A set (α) neighboring $(\alpha) = (0)$ determines through (2.2) two end points P and Q near the respective end points of g. If we assume for the moment that the end points of g are not conjugate, then P and Q can be joined by a unique extremal E, which is thus determined by the set (α). We can thus obtain a family of extremals determined by values of (α) near (0), and this family can be represented in the following form:

$$x_i = x_i^*(t, \alpha) \qquad (i = 1, \cdots, n), \tag{4.3}$$

where x_i^* and x_{it}^* are of class C'' in (α) and satisfy the following conditions:

$$x_i^*(t, 0) = \bar{x}_i(t) \qquad (i = 1, \cdots, n), \tag{4.4}$$

$$x_i^*(t^{(s)}, \alpha) = \overset{s}{x}_i(\alpha) \qquad (s = 1, 2). \tag{4.5}$$

The expression J taken along the extremals of the family (4.4) becomes a function $J(\alpha)$ of class C''.

The Euler equations (2.5) and the transversality conditions (2.6) enable us to prove that $J(\alpha)$ has a critical point for $(\alpha) = (0)$.

The terms of the second order of $J(\alpha)$ are obtained by means of the following identity in the variables $(u_1, \cdots, u_r)$:

$$J_{\alpha_h \alpha_k}(0) u_h u_k \equiv \frac{d^2 J}{de^2}(eu), \quad (e = 0) \qquad (h, k = 1, \cdots, r). \tag{4.6}$$

The right hand side of (4.6) is nothing but the second variation of the one-parameter family of extremals obtained from the family (4.3) by setting $\alpha_h = eu_h$, where u_h is fixed and e is variable. This one-parameter family has the form

$$x_i = x_i(t, e), \quad \alpha_h = eu_h \quad (i = 1, \cdots, n;\ h = 1, \cdots, r), \tag{4.7}$$

where

$$x_i(t^{(s)}, e) = x_i^s(eu) \qquad (i = 1, \cdots, n;\ s = 1, 2). \tag{4.8}$$

The second variation of the family (4.7) has the form (3.2), so that

$$J_{\alpha_h\alpha_k}(0)u_hu_k = b_{hk}u_hu_k + 2\int_{t^{(1)}}^{t^{(2)}} \omega(\eta, \dot\eta)dt \quad (h, k = 1, \cdots, r). \tag{4.9}$$

A curve $\eta_i = \eta_i(t)$ of class C'' in the space of the variables (t, η) will be called a *secondary extremal* if the functions (η) satisfy (3.7) for some σ. At present we are concerned only with secondary extremals for $\sigma = 0$.

To show the complete relation between (u) and (η) in (4.9), we need the following lemma.

LEMMA 2. *The integral $\int_{t_1}^{t_2} \omega dt$ has the same value if evaluated along any two secondary extremals joining the same end points $A:(t_1, a)$ and $B:(t_2, b)$.*

Suppose that $(\bar\eta)$ and $(\bar{\bar\eta})$ are the two secondary extremals. Then

$$\eta_i = \bar\eta_i + e(\bar\eta_i - \bar{\bar\eta}_i) \qquad (i = 1, \cdots, n)$$

is a one-parameter family of secondary extremals joining A and B and containing $(\bar\eta)$ and $(\bar{\bar\eta})$. But the value of an integral taken along the members of a one-parameter family of extremals joining the same end points is the same for each extremal.

Returning now to (4.9), we note that the functions $\eta_i(t)$ in the argument of ω define a secondary extremal E', since they are the variations of a family of extremals. The set (u) in (4.9) determines the end points of E'; for upon differentiating (4.8) with respect to e and setting $e = 0$, we obtain

$$\overset{s}{\eta}_i = \overset{s}{x}_{ih}u_h \quad (i = 1, \cdots, n;\ h = 1, \cdots, r;\ s = 1, 2), \tag{4.10}$$

and it is in this sense that the set (u) *determines* the end points of E'.

From (4.9) and Lemma 2 we obtain the following theorem:

THEOREM 4. *Under the assumption that the end points of g are not conjugate, let $J(\alpha)$ represent the value of J taken along the extremal determined by (α). Then the terms of second order of $J(\alpha)$ have the form*

$$J_{\alpha_h\alpha_k}(0)u_hu_k = b_{hk}u_hu_k + 2\int_{t^{(1)}}^{t^{(2)}} \omega(\eta, \dot\eta)dt \quad (h, k = 1, \cdots, r) \tag{4.11}$$

where (η) may be taken along any secondary extremal with end points determined by (u).

In order to bring the parameter σ into the second variation as in (3.5), we replace the integrand F by a one-parameter family of integrands

$$(4.12)\quad \bar{F} \equiv F - \frac{\sigma}{2}\{[x_i - \bar{x}_i(t)][x_i - \bar{x}_i(t)] + [\dot{x}_i - \bar{x}_i'(t)][\dot{x}_i - \bar{x}_i'(t)]\} \qquad (i = 1, \cdots, n)$$

which we consider only for $\sigma \leqq 0$. For $\sigma = 0$ we have our original problem in (x)-space, but for each $\sigma < 0$ we consider a non-parametric problem in (t, x)-space, the problem with the integral

$$\int_{t^{(1)}}^{t^{(2)}} \bar{F} dt$$

and the end conditions

$$(4.13)\qquad x_i^s = x_i^s(\alpha),\ t^s = t^{(s)} \qquad (i = 1, \cdots, n;\ s = 1, 2).$$

When we talk about extremals, conjugate points, etc., for $\sigma < 0$, these terms will always be understood to refer to the non-parametric problem in (t, x)-space.

For each $\sigma < 0$, $g: x_i = \bar{x}_i(t)$ is still an extremal. We note that the problem for each $\sigma < 0$ is *non-singular*; that is, along g the determinant

$$(4.14)\qquad \left|\frac{\partial^2 \bar{F}}{\partial \dot{x}_i \partial \dot{x}_j}\right| \equiv \left|\frac{\partial^2 F}{\partial \dot{x}_i \partial \dot{x}_j} - \sigma \delta_i^j\right| \neq 0, \qquad \sigma < 0.$$

This is a consequence of the Legendre sufficient condition. If, then, we assume momentarily that for each $\sigma < 0$ the end points of g are not conjugate, a set (α) neighboring (0) will determine for each $\sigma < 0$ a unique extremal, and the expression

$$\bar{J} = \int_{t^{(1)}}^{t^{(2)}} \bar{F} dt + \theta(\alpha)$$

becomes a function $\bar{J}(\alpha, \sigma)$. The following theorem is proved as was Theorem 4.

THEOREM 4a. *Under the assumption that the end points of g are not conjugate for any $\sigma \leqq 0$, let $\bar{J}(\alpha, \sigma)$ represent the value of $\bar{J}$ taken along the extremal determined by (α) for any $\sigma \leqq 0$. Then the terms of second order of $\bar{J}(\alpha, \sigma)$ have the form*

$$(4.15)\qquad H(u, \sigma) \equiv \bar{J}_{\alpha_h \alpha_k}(0, \sigma) u_h u_k = b_{hk} u_h u_k + 2 \int_{t^{(1)}}^{t^{(2)}} \Omega(\eta, \dot{\eta}, \sigma) dt \qquad (h, k = 1, \cdots, r).$$

For $\sigma < 0$, (η) is taken along the secondary extremal determined by (u) through (4.10), *while for $\sigma = 0$, (η) may be taken along any secondary extremal with end points determined by (u) through* (4.10).

5. **Sufficient conditions for a minimum.** Consider the expression

$$I(\eta, u, \sigma) = b_{hk}u_h u_k + 2\int_{t^{(1)}}^{t^{(2)}} \Omega(\eta, \dot\eta, \sigma)dt \quad (h, k = 1, \cdots, r).$$

By an admissible set (u, η) will be meant a set of constants (u) and a set of functions (η) of class D' which together satisfy (3.9).

THEOREM 5. *For sufficiently large negative values of σ, the expression $I(\eta, u, \sigma)$ is positive for all admissible sets $(u, \eta) \neq (0, 0)$.*

First we note that since $\|x_{ih}\|$ is of rank r, equations (3.9) can be solved for u_h in terms of a subset of the variables η_i^s. Hence for all admissible sets (u, η)

$$I(\eta, u, \sigma) = q(\eta) + 2\int_{t^{(1)}}^{t^{(2)}} \Omega(\eta, \dot\eta, \sigma)dt \tag{5.1}$$

where $q(\eta)$ is a form quadratic in the variables η_i^s. From this it follows† that for all admissible sets (u, η)

$$I(\eta, u, \sigma) \geqq \int_{t^{(1)}}^{t^{(2)}} [2\omega(\eta, \dot\eta) + M(\eta, \dot\eta) - \sigma\eta_i\eta_i - \sigma\dot\eta_i\dot\eta_i]dt \quad (i = 1, \cdots, n), \tag{5.2}$$

where $M(\eta, \dot\eta)$ is a suitably chosen form quadratic in the variables $(\eta, \dot\eta)$ with coefficient continuous in t.

But any such form as the integrand in (5.2) can be made positive definite by making σ negative and sufficiently large, independently of t.

Thus for such a σ,

$$I(\eta, u, \sigma) > 0 \tag{5.3}$$

for all admissible sets $(u, \eta) \neq (0, 0)$.

LEMMA 3. *Let (u, η) be any admissible set. If there is no point on g conjugate to its initial point for $\sigma = \sigma_0 < 0$, then*

$$H(u, \sigma_0) \leqq I(\eta, u, \sigma_0), \tag{5.4}$$

the equality holding if and only if (η) is a secondary extremal for $\sigma = \sigma_0$.

By Theorem 4a the equality holds if (η) is a secondary extremal for $\sigma = \sigma_0$. If (η) is not a secondary extremal, let $(\bar\eta)$ be the secondary extremal determined by (u) for $\sigma = \sigma_0$.

We note that along any arc (η) $(t^{(1)} \leqq t \leqq t^{(2)})$,

$$\frac{\partial^2\Omega}{\partial\dot\eta_i\partial\dot\eta_j}(\eta, \dot\eta, \sigma_0)\pi_i\pi_j = \frac{\partial^2\overline{F}}{\partial\dot x_i\partial\dot x_j}\pi_i\pi_j = \frac{\partial^2 F}{\partial\dot x_i\partial\dot x_j}\pi_i\pi_j - \sigma_0\pi_i\pi_i \quad (i, j = 1, \cdots, n),$$

† See Morse, p. 534, loc. cit.

which, by the Legendre condition, is positive for all $(\pi)\neq(0)$. Also, if we use Taylor's formula we see that the Weierstrass E-function

$$\Omega(\eta, \bar{\eta}', \sigma_0) - \Omega(\eta, \dot{\eta}, \sigma_0) - \frac{\partial\Omega}{\partial\dot{\eta}_i}(\eta, \dot{\eta}, \sigma_0)(\bar{\eta}_i' - \eta_i) \qquad (i = 1, \cdots, n)$$

is equal to

$$\frac{\partial^2\Omega}{\partial\dot{\eta}_i\partial\dot{\eta}_j}(\eta, \dot{\eta}, \sigma_0)(\bar{\eta}_i' - \dot{\eta}_i)(\bar{\eta}_j' - \dot{\eta}_j) \qquad (i, j = 1, \cdots, n)$$

and so is positive along any arc (η) and for all $(\bar{\eta}')\neq(\dot{\eta})$.

These facts, together with the hypothesis that there is no point on g conjugate to its initial point for $\sigma=\sigma_0$, enable us to infer that the secondary extremal $(\bar{\eta})$ minimizes $I(\eta, u, \sigma_0)$ in the fixed end point problem; that is, $I(\bar{\eta}, u, \sigma_0)<I(\eta, u, \sigma_0)$. The lemma follows from Theorem 4a.

LEMMA 4. *If $I(\eta, u, \sigma_0)$ $(\sigma_0<0)$ is positive for all admissible sets $(u, \eta)\neq(0,0)$, then there is no pair of conjugate points on g for $\sigma=\sigma_0$.*

For if t_2 were conjugate to t_1 on g for $\sigma=\sigma_0$, there would exist a secondary extremal $(\bar{\eta})\neq(0)$ vanishing at t_1 and t_2. Then $I(\eta, u, \sigma_0)$ would be zero if evaluated for $(u)=(0)$ and for (η) taken along the broken secondary extremal consisting of $(\bar{\eta})$ in the interval t_1t_2 and the t-axis in the remainder (if any) of the interval $t^{(1)}t^{(2)}$. This is contrary to hypothesis.

LEMMA 5. *If there is no point on g conjugate to its initial point for $\sigma=\sigma_0<0$, then there is no point on g conjugate to its initial point for σ in the neighborhood of σ_0.*

For each $\sigma<0$, the points conjugate to $t=t^{(1)}$ are defined by the zeros $t\neq t^{(1)}$ of the determinant $D(t, \sigma)=|\eta_{ij}(t, \sigma)|$, where $\|\eta_{ij}(t, \sigma)\|$ is a matrix each column of which represents a secondary extremal for $\sigma=\sigma$, and which satisfies the conditions

$$\|\eta_{ij}(t^{(1)}, \sigma)\| = \|0\|, \|\dot{\eta}_{ij}(t^{(1)}, \sigma)\| = \delta_i^j \ (i, j = 1, \cdots, n; \delta_i^j = \text{Kronecker delta}).$$

Now by means of the integral Law of the Mean, the function $\eta_{ij}(t, \sigma)$ can be expressed in the form

$$\begin{aligned}\eta_{ij}(t, \sigma) &= (t - t^{(1)})\int_0^1 \dot{\eta}_{ij}[t + \theta(t - t^{(1)}), \sigma]d\theta \ (i, j = 1, \cdots, n)\\ &= (t - t^{(1)})a_{ij}(t, \sigma),\end{aligned}$$

where $a_{ij}(t, \sigma)$ is continuous for $t^{(1)}\leqq t\leqq t^{(2)}$ and $\sigma<0$, and where

$$\|a_{ij}(t^{(1)}, \sigma)\| = \|\dot{\eta}_{ij}(t^{(1)}, \sigma)\| = \delta_i^j \qquad (i, j = 1, \cdots, n).$$

Thus

$$D(t, \sigma) = (t - t^{(1)})^n \left| a_{ij}(t, \sigma) \right| .$$

Since $D(t, \sigma_0) \neq 0$ for $t^{(1)} < t \leqq t^{(2)}$ by hypothesis, we see that $|a_{ij}(t, \sigma_0)| \neq 0$ for $t^{(1)} \leqq t \leqq t^{(2)}$. It follows from the continuity of $a_{ij}(t, \sigma)$ that $|a_{ij}(t, \sigma)|$ is $\neq 0$ in the interval $t^{(1)} \leqq t \leqq t^{(2)}$ for σ near σ_0. Hence $D(t, \sigma) \neq 0$ for σ near σ_0 in the interval $t^{(1)} < t \leqq t^{(2)}$, and the theorem is proved.

THEOREM 6. *If there exist no negative characteristic roots, then* $I(\eta, u, 0) \geqq 0$ *for all admissible sets* (u, η).

For σ negative and sufficiently large, $I(\eta, u, \sigma)$ is, by Theorem 5, positive for all admissible sets $(u, \eta) \neq (0, 0)$. Suppose we now increase σ towards zero. Then $I(\eta, u, \sigma)$ either remains positive for $\sigma < 0$ and for all admissible sets $(u, \eta) \neq (0, 0)$, or else there is a least upper bound $\sigma_0 < 0$ of the values of σ for which $I(\eta, u, \sigma)$ is positive for all admissible sets $(u, \eta) \neq (0, 0)$. We shall show that the latter case is impossible.

Suppose there does exist such a least upper bound σ_0. Then either $I(\eta, u, \sigma_0)$ is positive for all admissible sets $(u, \eta) \neq (0, 0)$, or else $I(\eta, u, \sigma_0)$ is zero for some such sets. If $I(\eta, u, \sigma_0)$ is zero for an admissible set $(\bar{u}, \bar{\eta}) \neq (0, 0)$ then $(\bar{u}, \bar{\eta})$ must minimize $I(\eta, u, \sigma_0)$ among admissible sets (u, η). Hence $(\bar{\eta})$ must be a secondary extremal for $\sigma = \sigma_0$ satisfying (3.8) and (3.9), contrary to the hypothesis that there exist no negative characteristic roots. Thus $I(\eta, u, \sigma_0)$ must be positive for all admissible sets $(u, \eta) \neq (0, 0)$.

Lemma 4 then enables us to set up the quadratic form $H(u, \sigma_0)$, which must be positive definite. By Lemma 5, we can set up $H(u, \sigma)$ for σ slightly greater than σ_0, and it must be positive definite for σ slightly greater than σ_0. By Lemma 3, $I(\eta, u, \sigma)$ must then be positive for all admissible sets $(u, \eta) \neq (0, 0)$ for σ slightly greater than σ_0. This contradicts the hypothesis that σ_0 is the least upper bound of the values of σ for which $I(\eta, u, \sigma)$ is positive for all admissible sets $(u, \eta) \neq (0, 0)$.

We conclude, then, that $I(\eta, u, \sigma)$ is positive for all $\sigma < 0$ and for all admissible sets $(u, \eta) \neq (0, 0)$.

It follows, then, that $I(\eta, u, 0) \geqq 0$ for all admissible sets (u, η).

A set of functions (η) will be called *tangential* if they are of the form

$$\eta_i = \rho(t) \bar{x}_i'(t) \tag{5.5}$$

where $\rho(t)$ is any function of class D'.

LEMMA 6. *A set of tangential functions of class* C'' *represents a secondary extremal for* $\sigma = 0$.

The one-parameter family

$$x_i = \bar{x}_i[t + e\rho(t)] \qquad (i = 1, \cdots, n),$$

where $\rho(t)$ is any function of class C'', is certainly a family of extremals, for its members are simply different representations of the same extremal g. Hence the variations $\eta_i(t)$ of this family represent a secondary extremal. But for this family

$$\eta_i = \bar{x}_i'(t)\rho(t) \qquad (i = 1, \cdots, n)$$

and so the lemma is proved.

Such a secondary extremal we shall call *tangential.*

LEMMA 7. *A tangential secondary extremal (not the t-axis) vanishing at $t^{(1)}$ and $t^{(2)}$ is a characteristic solution for $\sigma = 0$.*

That such a secondary extremal satisfies (3.8) for $\sigma = 0$, $(u) = (0)$, follows from the relation

$$\frac{\partial^2 F}{\partial \dot{x}_i \partial \dot{x}_j} \bar{x}_j' = 0 \qquad (i, j = 1, \cdots, n).$$

THEOREM 7. *If there exist no negative characteristic roots, and no non-tangential characteristic solutions for $\sigma = 0$, there is no point on g conjugate to its initial point for $\sigma = 0$.*

In the first place, $t^{(2)}$ cannot be conjugate to $t^{(1)}$ on g. For if it were, there would be a normal* secondary extremal $(\bar{\eta}) \not\equiv (0)$ vanishing at $t^{(2)}$ and $t^{(1)}$.† This curve $(\bar{\eta})$, with the set $(u) = (0)$, would make $I(\eta, u, 0)$ vanish. Now by Theorem 6, $I(\eta, u, 0)$ is positive or zero for all admissible sets (u, η) and so $(\bar{\eta})$ with the set $(u) = (0)$ would minimize $I(\eta, u, 0)$ among admissible sets (u, η). Hence $(\bar{\eta})$ would have to satisfy conditions (3.8) and so be a characteristic solution for $\sigma = 0$. Since $(\bar{\eta})$ is non-tangential, this is contrary to hypothesis.

Next suppose that $\bar{t} \neq t^{(2)}$ were conjugate to $t^{(1)}$ on g. Then there would exist a normal secondary extremal $(\bar{\eta}) \not\equiv (0)$ vanishing at $t^{(1)}$ and $\bar{t}$. The expression $I(\eta, u, 0)$ would be zero if evaluated along the broken secondary extremal (η) consisting of $(\bar{\eta})$ in the interval $t^{(1)}\bar{t}$ and of the t-axis in the remainder of the interval $t^{(1)}t^{(2)}$. The curve (η) would actually have a corner at $\bar{t}$, because the only normal secondary extremal through a point on the t-axis in the direction of the t-axis is $(\eta) \equiv (0)$.‡

---

* A *normal* secondary extremal is one which satisfies the relation

$$\bar{x}_i' \eta_i = 0 \qquad (i = 1, \cdots, n).$$

† See Bliss, loc. cit., p. 200.

‡ See Bliss, loc. cit., p. 199.

The arc (η) with the set $(u)=(0)$ would minimize $I(\eta, u, 0)$ among admissible sets (u, η), and so would have to satisfy the corner conditions

$$\left[\frac{\partial\omega}{\partial\dot{\eta}_i}\right]_{\bar{t}-}^{\bar{t}+} \equiv \frac{\partial^2 F}{\partial\dot{x}_i\partial\dot{x}_j}[\dot{\eta}_j]_{\bar{t}-}^{\bar{t}+} = 0 \qquad (i, j = 1, \cdots, n).$$

From this it would follow, due to the actual presence of a corner at $\bar{t}$, that

$$[\dot{\eta}_j]_{\bar{t}-}^{\bar{t}+} = k\bar{x}_j'(t) \qquad k \neq 0 \quad (j = 1, \cdots, n). \tag{5.6}$$

Hence

$$\bar{\eta}_j'(\bar{t}) = -k\bar{x}_j'(\bar{t}) \qquad (j = 1, \cdots, n). \tag{5.7}$$

But this is impossible; for along the normal secondary extremal $(\bar{\eta})$ we have

$$\bar{x}_j'\bar{\eta}_j = 0 \qquad (j = 1, \cdots, n), \tag{5.8}$$

and hence, by differentiation,

$$\bar{x}_j'(\bar{t})\bar{\eta}_j'(\bar{t}) = 0 \qquad (j = 1, \cdots, n), \tag{5.9}$$

which contradicts (5.7).

Thus there is no point on g conjugate to $t^{(1)}$.

We come now to the final theorem. The arc g and the set $(\alpha)=(0)$ shall be said to furnish a proper, strong, relative minimum to J if there exist a neighborhood N of g and a neighborhood M of $(\alpha)=(0)$ such that the value of J is less when evaluated for g and $(\alpha)=(0)$ than when evaluated for any other admissible arc in N with ends determined by a set (α) in M.

THEOREM 8. *In order that the extremal g, without multiple points, and the set $(\alpha)=(0)$ afford a proper strong relative minimum to J it is sufficient that the transversality conditions* (2.6) *be satisfied, that the Legendre and Weierstrass sufficient conditions hold, that there be no negative characteristic roots, and that there be no characteristic solutions for $\sigma=0$ except the tangential solutions vanishing at both ends.*

Under the hypotheses of this theorem, Theorem 7 tells us that the end points of g are not conjugate, and so we can set up the function $\bar{J}(\alpha, 0)$, and hence the quadratic form $H(u, 0)$. According to Theorem 4, $H(u, 0)$ is equal to $I(\eta, u, 0)$, where (η) is any secondary extremal with ends determined by (u) through (3.9). By Theorem 6, $H(u, 0) \geq 0$.

Now if $H(u, 0)$ were 0 for some $(u)\neq(0)$, then $I(\eta, u, 0)$ would be zero if evaluated for (u) and any secondary extremal $(\bar{\eta})$ with ends determined by (u). Hence $(\bar{\eta})$ would minimize $I(\eta, u, 0)$ and so would satisfy (3.8) and be a characteristic solution for $\sigma=0$ not vanishing at both ends. This contradicts the hypotheses. Thus $H(u\ 0)$ is positive definite.

Now the Legendre and Weierstrass sufficient conditions are assumed to hold along g. Also, by Theorem 7, there is no point on g conjugate to its initial point. Hence g furnishes a minimum to J in the fixed end point problem. Furthermore, there exists a neighborhood N of g such that if an extremal E determined by a set (α) lies in N, then, if (α) is sufficiently near (0), E will afford a minimum to J in the fixed end point problem, with respect to admissible arcs in N joining the end points of E.*

Let g' be any admissible arc in N, its end points being given by a certain set (α). Then if (α) is near enough to (0) the extremal determined by (α) will lie in N, and

$$J_{g',\alpha} \geqq J(\alpha). \tag{5.10}$$

But $H(\alpha, 0)$ gives the terms of second order in $J(\alpha)$, so that for (α) sufficiently near (0) we have

$$J(\alpha) \geqq J(0), \tag{5.11}$$

the equality holding only if $(\alpha) = (0)$.

Hence for g' sufficiently near g and (α) sufficiently near (0),

$$J_{g',\alpha} \geqq J(0). \tag{5.12}$$

This inequality becomes an equality only if g' is identical with g.

Thus the theorem is proved.

6. **The fixed end point problem.** This is the case that $r=0$ and $\theta=$constant, the end conditions being

$$x_i^s = \text{constant} \qquad (i = 1, \cdots, n;\ s = 1, 2).$$

The expression $I(\eta, u, \sigma)$ is replaced by

$$I(\eta, \sigma) = 2\int_{t^{(1)}}^{t^{(2)}} \Omega dt,$$

and the accessory boundary problem has the form

$$\frac{d}{dt}\left[\frac{\partial \Omega}{\partial \dot{\eta}_i}\right] - \frac{\partial \Omega}{\partial \eta_i} = 0 \qquad (i = 1, \cdots, n),$$

$$\eta_i^s = 0 \qquad (i = 1, \cdots, n;\ s = 1, 2).$$

The necessary condition of Theorem 3 holds as stated.

To prove Theorem 8 in the fixed end point case, we shall prove that under the hypotheses of the theorem there is no point on g conjugate to its

* Cf. Morse, loc. cit., p. 535, and Bliss, Annals of Mathematics, April, 1932, p. 267, Lemma 1.

initial point for $\sigma=0$. This will follow if we can prove Theorem 7, which in turn is based on Theorem 6.

The first two paragraphs in the proof of Theorem 6 hold as before. Next, Lemma 4 shows us that there is no point on g conjugate to its initial point for $\sigma=\sigma_0$, and Lemma 5 extends this property to values of σ slightly greater than σ_0. Hence $(\eta)\equiv(0)$ furnishes a proper minimum to $I(\eta, \sigma)$ (see proof of Lemma 3) for these values of σ, and so $I(\eta, \sigma)>0$ for these values of σ for $(\eta)\not\equiv(0)$. This contradicts the hypothesis that σ_0 was the least upper bound of the values of σ for which $I(\eta, \sigma)$ is positive for all admissible sets $(\eta)\not\equiv(0)$. Theorem 6 follows, and hence Theorems 7 and 8.

HARVARD UNIVERSITY,
CAMBRIDGE, MASS.

NOTES ON THE THEORY AND APPLICATION OF FOURIER TRANSFORMS. III, IV, V, VI, VII*

BY

R. E. A. C. PALEY AND N. WIENER

III. On Müntz's theorem

1. We give a proof of the following theorem which is Szász's† generalization of Müntz's‡ theorem. The method is similar to one employed by Carleman§ to prove Müntz's theorem. Incidentally we give a theorem concerning the distribution of the zeros of functions analytic in a half-plane, which is analogous to, and in some respects more general than, another theorem of Carleman's paper.

2. We recall the following well known

Theorem I. *Let*

$$\zeta_n = \rho_n e^{i\theta_n},\ 0 \leqq \rho_n < 1 \qquad (n = 1, 2, \cdots) \tag{2.1}$$

be a set of points in the unit circle $|\zeta| < 1$. *Suppose that an analytic function* $f_1(\zeta)$ *is regular in* $|\zeta| < 1$, *satisfies*

$$\int_{-\pi}^{\pi} |f_1(\rho e^{i\theta})|^2 d\theta \leqq B,\ \rho < 1, \tag{2.2}$$

where B is a constant which depends only on f_1, *and has zeros at the points* ζ_n. *Then*

$$\sum_{n=1}^{\infty}(1 - \rho_n) < \infty. \tag{2.3}$$

Conversely, if the series (2.3) *converges, then there exists a bounded function* $f_1(\zeta)$ *which has zeros at the points* ζ_n.

Suppose we invert the interior of the unit circle into the half-plane $\Im(z) > 0$, by means of the substitution

* Presented to the Society, October 28, 1933; received by the editors March 16, 1933. Notes I and II of the Series have appeared in this volume of these Transactions, pp. 348–355.

† O. Szász, *Über die Approximation stetiger Funktionen durch lineare Aggregate von Potenzen*, Mathematische Annalen, vol. 77 (1916), pp. 482–496.

‡ C. H. Müntz, *Über den Approximationssatz von Weierstrass*, Schwarz's Festschrift, Berlin, 1914, pp. 303–312.

§ T. Carleman, *Über die Approximation analytischer Funktionen durch lineare Aggregate von vorgegebenen Potenzen*, Arkiv för Matematik, Astronomi och Fysik, vol. 17 (1922–23), No. 9, pp. 1–30.

$$\zeta = \frac{1 + iz}{1 - iz}. \tag{2.4}$$

Suppose that the point $\zeta = \rho e^{i\theta}$ inverts into $z = x + iy$. Then it is a matter of elementary algebra to verify that the convergence of the series (2.3) is equivalent to that of the series

$$\sum_{n=1}^{\infty} \frac{y_n}{1 + x_n^2 + y_n^2},$$

$z_n = x_n + iy_n$, $y_n > 0$, being the inverse of $\zeta_n = \rho_n e^{i\theta_n}$.

We are now in a position to prove our next theorem, which is a generalization of Carleman's:

THEOREM II. *Let*

$$z_n = x_n + iy_n \qquad (n = 1, 2, \cdots) \tag{2.5}$$

denote a sequence of points in the half plane $\Im(z) > 0$, *and let* $f(z)$ *be a function regular in* $\Im(z) > 0$ *which satisfies*

$$\int_{-\infty}^{\infty} |f(x + iy)|^2 dx < 1 \tag{2.6}$$

and has zeros at the points z_n. *Then*

$$\sum_{n=1}^{\infty} \frac{y_n}{1 + x_n^2 + y_n^2} < \infty. \tag{2.7}$$

Conversely, if the series (2.7) *converges, then there exists a bounded function* $f(z)$ *regular in* $\Im(z) > 0$, *satisfying* (2.6) *and having zeros at the points* z_n.

For the second half of the theorem we have only to observe, by Theorem I, that there exists a function $g(z)$ with zeros at the points z_n, analytic in $\Im(z) > 0$, and less than 1 in absolute value. We have now only to write

$$f(z) = \tfrac{1}{2}(z + i)^{-1} g(z)$$

and condition (2.6) is satisfied.

To prove the first part of the Theorem it will be sufficient to show that the function $f(z)$ can be represented by a Cauchy integral

$$f(z) = (2\pi i)^{-1} \int_{-\infty}^{\infty} \frac{f^*(x')}{x' - z} dx', \ \Im(z) > 0, \tag{2.8}$$

where the function $f^*(x')$ satisfies

$$\int_{-\infty}^{\infty} |f^*(x')|^2 dx' \leqq 1. \tag{2.9}$$

Indeed, it is readily seen that the substitution (2.4) transforms the integral of the right-hand member of (2.8) into

$$(2\pi i)^{-1}\int_{-\pi}^{\pi}\frac{f_1^*(\zeta')d\zeta'}{\zeta'-\zeta}+C,\ \zeta'=e^{i\phi},$$

where $f_1^*(\zeta')$ corresponds to $f^*(x')$ and

$$C=(2\pi i)^{-1}\int_{-\infty}^{\infty}\frac{f^*(x')}{x'+i}dx'.$$

It follows that, on putting $f_1(\zeta)\equiv f(z)$,

$$\int_{-\pi}^{\pi}|f_1(\rho e^{i\theta})|^2dt\leqq B$$

and we may apply Theorem I.

Now suppose that

$$0<\epsilon<y<y_0,\quad 2|x|<N_0<N.$$

Then

$$2\pi if(z)=\int_{-N}^{N}\frac{f(x'+i\epsilon)}{x'+i\epsilon-z}dx'-\int_{-N}^{N}\frac{f(x'+iy_0)}{x'+iy-z}dx'$$
$$+i\int_{\epsilon}^{y_0}\frac{f(N+iy')}{N+iy'-z}dy'-i\int_{\epsilon}^{y_0}\frac{f(-N+iy')}{-N+iy'-z}dy'.$$

But

$$\frac{1}{N_0}\int_{N_0}^{2N_0}dN\int_{\epsilon}^{y_0}\frac{|f(N+iy')|}{|N+iy'-z|}dy'\leqq\frac{2}{N_0^2}\int_{\epsilon}^{y_0}dy'\int_{N_0}^{2N_0}|f(N+iy')|\,dN$$
$$\leqq\frac{2}{N_0^2}\int_{\epsilon}^{y_0}dy'N_0^{1/2}\left[\int_{N_0}^{2N_0}|f(N+iy')|^2dN\right]^{1/2}\leqq 2(y_0-\epsilon)N_0^{-3/2}$$

and tends to zero as $N_0\to\infty$. Similar analysis applies to the last term. Hence

$$2\pi if(z)=\lim_{N_0\to\infty}\frac{1}{N_0}\int_{N_0}^{2N_0}dN\left[\int_{-N}^{N}\frac{f(x'+i\epsilon)}{x'+i\epsilon-z}dx'-\int_{-N}^{N}\frac{f(x'+iy_0)}{x'+iy_0-z}dx'\right]$$
$$=\int_{-\infty}^{\infty}\frac{f(x'+i\epsilon)}{x'+i\epsilon-z}dx'-\int_{-\infty}^{\infty}\frac{f(x'+iy_0)}{x'+iy_0-z}dx',$$

since the last two integrals converge absolutely. Now make y_0 tend to infinity. The second integral does not exceed

$$\left[\int_{-\infty}^{\infty} |f(x'+iy_0)|^2 dx'\right]^{1/2} \left[\int_{-\infty}^{\infty} \frac{dx'}{|x'+iy_0-z|^2}\right]^{1/2},$$

and so tends to zero.

Since condition (2.6) is satisfied, by a classical argument of F. Riesz† there exists a sequence $\{\epsilon_k \downarrow 0\}$ such that $f(x+i\epsilon_k)$ converges weakly to a function $f^*(x)$ satisfying (2.9). Hence

$$2\pi i f(z) = \int_{-\infty}^{\infty} \frac{f(x'+i\epsilon_k)}{x'-z} dx' \to \int_{-\infty}^{\infty} \frac{f^*(x')}{x'-z} dx' \text{ as } \epsilon_k \to 0,$$

which is the desired result.

3. We proceed now to our main theorem.

THEOREM III. *A necessary and sufficient condition for the closure* L^2 *of the functions* t^{λ_n}, $\Re(\lambda_n) > -\frac{1}{2}$, *in the interval* (0, 1), *is the divergence of the series*

$$\sum_{n=1}^{\infty} \frac{1+2\Re(\lambda_n)}{1+|\lambda_n|^2}. \tag{3.1}$$

We observe first that if the functions t^{λ_n} are not closed L^2 then there exists a function $\phi(t)$ of integrable square on (0, 1) which is orthogonal to them all, so that we have

$$0 < \int_0^1 |\phi(t)|^2 dt < \infty, \tag{3.2}$$

$$\int_0^1 \phi(t) t^{\bar{\lambda}_n} dt = 0 \qquad (n = 1, 2, 3, \cdots). \tag{3.3}$$

Conversely if the system $\{t^{\lambda_n}\}$ is closed then there exists no function $\phi(t)$ not identically zero satisfying (3.2) and (3.3). Now let us write $t = e^x$; the conditions (3.2) and (3.3) become

$$0 < \int_{-\infty}^{0} |\phi(e^x)|^2 e^x dx < \infty,$$

$$\int_{-\infty}^{0} \phi(e^x) \exp\left[x(1+\bar{\lambda}_n)\right] dx = 0 \qquad (n = 1, 2, \cdots).$$

Upon writing $\phi(e^x)e^{x/2} = \Phi(x)$, we transform the last two formulas into

† *Untersuchungen über Systeme integrierbarer Funktionen*, Mathematische Annalen, vol. 69 (1910), pp. 449–497; pp. 466–468.

$$0 < \int_{-\infty}^{0} |\Phi(x)|^2 dx < \infty, \tag{3.4}$$

$$\int_{-\infty}^{0} \Phi(x) \exp [x(\tfrac{1}{2} + \bar{\lambda}_n)] dx = 0 \qquad (n = 1, 2, \cdots). \tag{3.5}$$

Thus the closure or non-closure L^2 of the functions t^{λ_n} on (0, 1) is equivalent to the non-existence or existence of a function satisfying (3.4) and (3.5), which is equivalent to the closure or non-closure of the functions $\exp [x(\frac{1}{2}+\lambda_n)]$ on the interval $(-\infty, 0)$.

4. Proceeding now to the proof of the theorem we observe first that, if there exists a function satisfying (3.4) and (3.5), then the function

$$f(z) = \int_{-\infty}^{0} \Phi(x') e^{-izx'} dx$$

exists for $\mathfrak{J}(z) > 0$, and defines an analytic function in that half-plane. Further, by (3.5), it vanishes at the points $(\frac{1}{2}+\bar{\lambda}_n)i$. Since, by Plancherel's theorem,

$$\int_{-\infty}^{\infty} |f(x + iy)|^2 dx = 2\pi \int_{-\infty}^{0} |\Phi(x')|^2 e^{2yx'} dx',$$

the series (3.1) converges by Theorem II. Thus the non-closure of the functions $\{x^{\lambda_n}\}$ implies the convergence of (3.1).

To obtain the converse we have to show that when (3.1) converges then we can find a function $\Phi(x)$ satisfying (3.4) and (3.5). Now in virtue of Theorem II we can find a function $f(z)$ which is analytic for $\mathfrak{J}(z) > 0$, is uniformly bounded in this half-plane, and vanishes at the points $(\frac{1}{2}+\bar{\lambda}_n)i$, with the integral

$$\int_{-\infty}^{\infty} |f(x + iy)|^2 dx$$

uniformly bounded. Let $g_y(\xi)$ denote the Fourier transform

$$(2\pi)^{-1/2} \int_{-\infty}^{\infty} f(x + iy) e^{ix\xi} dx$$

of $f(x+iy)$. Then the argument given in detail in the first note of this series (Theorem II) shows that $g_y(\xi)$ is of the form $G(\xi)e^{y\xi}$ for $\xi < 0$ and vanishes for $\xi > 0$, where

$$\int_{-\infty}^{0} |G(\xi)|^2 d\xi < \infty .$$

Now

$$(2\pi)^{-1/2}\int_{-\infty}^{0} G(\xi)\exp\left[\xi(\tfrac{1}{2}+\bar{\lambda}_n)\right]d\xi$$

$$= (2\pi)^{-1/2}\int_{-\infty}^{0} G(\xi)\exp\left[\xi\Re(\tfrac{1}{2}+\lambda_n)\right]\exp\left[-i\xi\Im(\lambda_n)\right]d\xi$$

$$= (2\pi)^{-1/2}\int_{-\infty}^{0} g_y(\xi)\exp\left[-i\xi\Im(\lambda_n)\right]d\xi \qquad (y = \Re(\tfrac{1}{2}+\lambda_n))$$

$$= f\{\Im(\lambda_n) + i\Re(\tfrac{1}{2}+\lambda_n)\} = f\{(\tfrac{1}{2}+\bar{\lambda}_n)i\} = 0.$$

Thus, for all n,

$$\int_{-\infty}^{0} G(\xi)\exp\left[\xi(\tfrac{1}{2}+\bar{\lambda}_n)\right]d\xi = 0.$$

We have only to identify $\Phi(x)$ with $G(x)$ and our theorem is proved.

5. The problem of the closure in L^2 of functions $e^{\lambda_n x}$ on a finite interval is much more difficult than the corresponding one for an interval which is infinite in one direction. We have obtained a number of theorems in this direction but nothing like a complete answer to the problem. In the case however where all the numbers λ_n are real (we need no longer assume that λ_n is positive or negative) a necessary and sufficient condition for the closure of the functions $e^{\lambda_n x}$ on a finite interval is the divergence of the series

$$\sum_{n=1}^{\infty}\frac{1}{1+|\lambda_n|}.$$

IV. A theorem on closure

1. The present note is devoted to the proof of the following theorem:

Theorem I. *The set of functions* $\{e^{-\pi|x|/2}e^{i\lambda_n x}\}$ *is closed* L^2 *over* $(-\infty,\infty)$ *when and only when*

$$\sum_{n=1}^{\infty}\frac{\cos\Im(\lambda_n)}{\cosh\Re(\lambda_n)} = \infty, \quad -\tfrac{1}{2}\pi < \Im(\lambda_n) < \tfrac{1}{2}\pi. \tag{1.1}$$

The condition (1.1) should be contrasted with the condition

$$\sum_{n=1}^{\infty}\frac{\Im(\lambda_n)-\frac{1}{2}\pi}{1+|\lambda_n|^2} = \infty, \quad \Im(\lambda_n) > -\tfrac{1}{2}\pi,$$

which is a necessary and sufficient condition for the closure L^2 of the functions $\{e^{-\pi|x|/2}e^{i\lambda_n x}\}$ on the interval $(0,\infty)$ (see e.g. the preceding note of this series). Thus if for instance all the numbers λ_n are real, the conditions for closure on the intervals $(0,\infty)$ and $(-\infty,\infty)$ are

$$\sum_{n=1}^{\infty} \frac{1}{1+|\lambda_n|^2} = \infty, \quad \sum_{n=1}^{\infty} e^{-|\lambda_n|} = \infty$$

respectively.

2. We shall prove the theorem by the following chain of lemmas:

LEMMA 1. *As $|x| \to \infty$ in either direction along the real axis,*

$$|\Gamma(ix+\tfrac{1}{2})| \sim (2\pi)^{1/2} e^{-\pi|x|/2}.$$

This is an immediate consequence of Stirling's formula.

LEMMA 2. *The set of functions $\{e^{-\pi|x|/2} e^{i\lambda_n x}\}$ is closed L^2 when and only when the set $\{\Gamma(ix+\frac{1}{2})e^{i\lambda_n x}\}$ is closed L^2.*

For let

$$\int_{-\infty}^{\infty} |f(x)|^2 dx < \infty \tag{2.1}$$

and

$$\int_{-\infty}^{\infty} f(x)\Gamma(ix+\tfrac{1}{2})e^{i\lambda_n x} dx = 0 \qquad (n = 1, 2, \cdots). \tag{2.2}$$

Let

$$g(x) = f(x)\Gamma(ix+\tfrac{1}{2})e^{\pi|x|/2}.$$

Then, by Lemma 1,

$$\int_{-\infty}^{\infty} |g(x)|^2 dx < \infty \tag{2.3}$$

and

$$\int_{-\infty}^{\infty} g(x)e^{-\pi|x|/2} e^{i\lambda_n x} dx = 0 \qquad (n = 1, 2, \cdots). \tag{2.4}$$

Similarly, (2.1) and (2.2) follow from (2.3) and (2.4). Thus there exists a function $f(x)$ orthogonal to the set $\{\Gamma(ix+\frac{1}{2})e^{i\lambda_n x}\}$ when and only when there exists a function $g(x)$ orthogonal to the set $\{e^{-\pi|x|/2}e^{i\lambda_n x}\}$, and one set is closed L^2 when and only when the other set is closed.

LEMMA 3. *The set of functions*

$$e^{v/2} \exp(-e^{v-\lambda_n}) \qquad (n = 1, 2, \cdots) \tag{2.5}$$

is closed L^2 over $(-\infty, \infty)$ when and only when the set of functions $\{\Gamma(ix+\frac{1}{2})e^{i\lambda_n x}\}$ is closed L^2 over $(-\infty, \infty)$.

This follows from the fact that the Fourier transform of (2.5), apart from a constant factor, is $\Gamma(ix+\frac{1}{2})e^{i\lambda_n x}$, and that, by Plancherel's theorem, L^2-closure is invariant under a Fourier transformation.

LEMMA 4. *The set of functions* $\{x^{\mu_n}\}$, $\mu_n=e^{-\lambda_n}-\frac{1}{2}$, *is closed* L^2 *over* (0, 1) *when and only when the set of functions* (2.5) *is closed* L^2 *over* $(-\infty, \infty)$.

For the pair of statements

$$\int_0^1 |f(x)|^2 dx < \infty$$

and

$$\int_0^1 f(x) x^{\mu_n} dx = 0$$

is equivalent to the pair of statements

$$\int_{-\infty}^{\infty} |g(v)|^2 dv = 0$$

and

$$\int_{-\infty}^{\infty} g(v) e^{v/2} \exp(-e^{v-\lambda_n}) dv = 0,$$

where

$$g(v) = f[\exp(-e^v)] e^{v/2} \exp(-e^v/2),$$

and we may again apply the argument of Lemma 2.

LEMMA 5. *The set of functions* $\{x^{\mu_n}\}$, $\Re(\mu_n) > -\frac{1}{2}$, *is closed* L^2 *over* (0, 1) *when and only when*

$$\sum_{n=1}^{\infty} \frac{2\Re(\mu_n)+1}{1+|\mu_n|^2} = \infty . \tag{2.6}$$

This is a well known theorem of Szász.†

Condition (2.6) is equivalent to (1.1). Theorem I now follows by combining Lemmas 2, 3, 4, and 5.

V. ON ENTIRE FUNCTIONS

1. Let $f(z)$ be an entire function, $f(0)=1$, and $\{z_\nu\}$ the sequence of zeros of $f(z)$. We denote by $M_f(r)$ and $m_f(r)$ respectively the maximum and minimum of $|f(z)|$ on the circle $|z|=r$, and by $n_f(r)$ the number of zeros of $f(z)$

† Loc. cit. See also our preceding note.

contained in the interior of this circle. The purpose of this note is to prove the following two theorems:

THEOREM I. *Let*

$$\log M_f(r) = O(r^{1/2}) \tag{1.1}$$

and

$$\int_0^\infty \log^+ m_f(r) r^{-3/2} dr < \infty . \tag{1.2}$$

Then

$$n_f(r) \sim A r^{1/2} \tag{1.3}$$

where the constant A is determined by

$$A = -\pi^{-2} \int_0^\infty \log \prod_{\nu=1}^{\infty} \left| \left(1 - \frac{x}{|z_\nu|}\right) \right| x^{-3/2} dx. \tag{1.4}$$

THEOREM II. *Let $f(z)$ be an entire function of order not exceeding $\frac{1}{2}$. If the conditions*

$$n_f(r) \sim B r^{1/2}, \tag{1.5}$$

$$B = -\pi^{-2} \int_0^\infty \log | f(x) | x^{-3/2} dx, \quad f(0) = 1, \tag{1.6}$$

are satisfied, then all roots of $f(z)$ are positive.

The proofs of these theorems are based upon a lemma which is of independent interest. This lemma is discussed in the next §2. In §3 we give proofs of Theorems I and II. In §4 we give a modification of the lemma of §2 and discuss its application to the theory of the Riemann zeta-function. In the last §5 we give proof of some results analogous to those of §2. They are in part contained in a paper by Titchmarsh,† and in part represent extensions of his results.

2. Let $\{\lambda_\nu\}$ be a monotone sequence of positive numbers such that the series $\sum_1^\infty \lambda_\nu^{-2}$ converges. We set

$$\phi(z) = \prod_{\nu=1}^{\infty} \left(1 - \frac{z^2}{\lambda_\nu^2}\right). \tag{2.1}$$

† E. C. Titchmarsh, *On integral functions with real negative zeros*, Proceedings of the London Mathematical Society, (2), vol. 26 (1927), pp. 185–200.

LEMMA 2.1. *If* $\sum_1^\infty \lambda_\nu^{-2}$ *converges then the statements*

$$\log \phi(iy) \sim \pi A\,|y| \text{ as } |y| \to \infty, \tag{2.2}$$

$$\int_{-\infty}^{\infty} \log|\phi(x)|\, x^{-2}dx = -\pi^2 A, \tag{2.3}$$

are completely equivalent.

We have, assuming $y>0$,

$$\begin{aligned}(\pi y)^{-1}\log\phi(iy) &= (\pi y)^{-1}\sum_{\nu=1}^{\infty}\log\left(1+\frac{y^2}{\lambda_\nu^2}\right)\\ &= (\pi y)^{-1}\int_0^\infty \log\left(1+\frac{y^2}{t^2}\right)d\Lambda(t)\end{aligned} \tag{2.4}$$

where $\Lambda(t)$ is the number of λ_ν's not exceeding t. Similarly

$$\begin{aligned}-\pi^{-2}\int_{-v}^{v}\log|\phi(x)|\,x^{-2}dx &= -2\pi^{-2}\int_0^v x^{-2}dx\int_0^\infty \log\left|1-\frac{x^2}{t^2}\right|d\Lambda(t)\\ &= -2\pi^{-2}\int_0^\infty d\Lambda(t)\int_0^v \log\left|1-\frac{x^2}{t^2}\right|x^{-2}dx\\ &= -2\pi^{-2}y^{-1}\int_0^\infty d\Lambda(t)\frac{y}{t}\int_0^{v/t}\log|1-s^2|\,s^{-2}ds.\end{aligned} \tag{2.5}$$

Expressions (2.4), (2.5) are both of the form

$$\frac{1}{y}\int_0^\infty N\left(\frac{t}{y}\right)d\Lambda(t) \tag{2.6}$$

where $\Lambda(t)$ is a monotone increasing function. In (2.4) we have

$$N(\lambda) = N_1(\lambda) = \frac{1}{\pi}\log\left(1+\frac{1}{\lambda^2}\right), \tag{2.7}$$

while, in (2.5),

$$\begin{aligned}N(\lambda) = N_2(\lambda) &= -\frac{2}{\pi^2\lambda}\int_0^{1/\lambda}\log|1-x^2|\,x^{-2}dx\\ &= \frac{2}{\pi^2}\left\{\log\left|1-\frac{1}{\lambda^2}\right| + \frac{1}{\lambda}\log\left|\frac{1+\lambda}{1-\lambda}\right|\right\}.\end{aligned} \tag{2.8}$$

The function $N_1(\lambda)$ is positive and monotone decreasing, the same being also true of $N_2(\lambda)$ since

$$N_2'(\lambda) = -\frac{2}{\pi^2\lambda^2}\log\left|\frac{1+\lambda}{1-\lambda}\right|. \tag{2.9}$$

If we write $N(\lambda)$ for either of $N_1(\lambda)$, $N_2(\lambda)$, the following properties are easily established:

$$N(\lambda) = \begin{cases} O\left(\log\dfrac{1}{\lambda}\right) & \text{as } \lambda \to 0, \\ O\left(\dfrac{1}{\lambda^2}\right) & \text{as } \lambda \to \infty, \end{cases} \tag{2.10}$$

$$\sum_{k=-\infty}^{\infty} \max_{2^k \le \lambda \le 2^{k+1}} \lambda N(\lambda) < \infty, \; N(\lambda) > 0, \tag{2.11}$$

$$\int_0^\infty N_1(\lambda)\lambda^{it}d\lambda = \frac{1}{\pi(it+1)}\int_0^\infty \frac{\mu^{(it-1)/2}d\mu}{1+\mu} = \frac{1}{(it+1)\cosh\dfrac{\pi t}{2}},$$

$$\int_0^\infty N_2(\lambda)\lambda^{it}d\lambda = \frac{2}{\pi^2(it+1)}\int_0^\infty \lambda^{it-1}\log\left|\frac{1+\lambda}{1-\lambda}\right|d\lambda$$

$$= \frac{2}{\pi^2(it+1)}\sum_{k=-\infty}^{\infty}\frac{1}{\left(\dfrac{t}{2}\right)^2 + \left(k+\dfrac{1}{2}\right)^2} = \frac{2\tan\dfrac{\pi it}{2}}{\pi it(it+1)}.$$

It follows that

$$\int_0^\infty N(\lambda)\lambda^{it}d\lambda \ne 0 \text{ when } t \text{ is real}, \tag{2.12}$$

$$\int_0^\infty N(\lambda)d\lambda = 1. \tag{2.13}$$

We observe finally that the expressions

$$\frac{1}{y}\int_0^\infty N_1\left(\frac{t}{y}\right)d\Lambda(t) = (\pi y)^{-1}\log\phi(iy),$$

$$\frac{1}{y}\int_0^\infty N_2\left(\frac{t}{y}\right)d\Lambda(t) = -\pi^{-2}\int_{-y}^{y}\log|\phi(x)|\,x^{-2}dx$$

$\to 0$ as $y \to 0$. Hence either of the statements

$$\frac{1}{y}\int_0^\infty N_i\left(\frac{t}{y}\right)d\Lambda(t) \to A \text{ as } y \to \infty,\; i = 1, 2, \tag{2.14}$$

implies the boundedness of the corresponding integral

$$\frac{1}{y}\int_0^\infty N_i\left(\frac{t}{y}\right)d\Lambda(t)$$

over the range $(0, \infty)$. A direct application of a Tauberian theorem of Wiener† shows that statements (2.14) are completely equivalent, which is precisely the result of our Lemma 2.1.

3. We now proceed to the proofs of Theorems 1 and 2.

Proof of Theorem I. We first observe that by a known theorem the assertions $\log M_f(r) = O(r^{1/2})$ and $n_f(r) = O(r^{1/2})$ are equivalent. It follows that if we replace each zero by another zero with the same absolute value but situated on the positive part of the real axis, changing in effect

$$\prod_1^\infty\left(1-\frac{z}{z_\nu}\right) \text{ into } \prod_1^\infty\left(1-\frac{z}{|z_\nu|}\right),$$

we certainly do not affect the truth of (1.1). Secondly this process will decrease $m_f(r)$ for every value of r, so that we do not affect the truth of (1.2) either. Thus it is legitimate to assume that all the zeros of $f(z)$ are real and positive. Let them be

$$\lambda_1^2, \lambda_2^2, \cdots; \quad 0 < \lambda_1 \leqq \lambda_2 \leqq \cdots; \quad \sum_1^\infty \lambda_\nu^{-2} < \infty .$$

Our next observation is that, by some theorems of Titchmarsh‡, for a function of this special type the assertions

$$n_f(r) \sim Ar^{1/2} \text{ and } \log M_f(r) \sim \pi A r^{1/2}$$

are equivalent, so that we may leave $n_f(r)$ and confine our attention to $M_f(r)$. We write

$$z^{1/2} = w = u + iv$$

so that $f(z)$ is transformed into the new function

$$f(w^2) = \phi(w) = \prod_{\nu=1}^\infty\left(1-\frac{w^2}{\lambda_\nu^2}\right)$$

which satisfies

$$\log^+|\phi(w)| = O(|w|) \tag{3.1}$$

and

† N. Wiener, *Tauberian theorems*, Annals of Mathematics, (2), vol. 33 (1932), pp. 1–100; Theorem XI′, p. 30.

‡ Loc. cit., Theorems I and II.

$$\int_{-\infty}^{\infty} \log^+ |\phi(u)|\, u^{-2} du < \infty . \tag{3.2}$$

We have to show that

$$\log \phi(iv) \sim \pi A\, |v| . \tag{3.3}$$

Since the series $\sum_1^\infty \lambda_\nu^{-2}$ converges, by Lemma 2.1 it will be sufficient to establish the convergence of the integral

$$\int_{-\infty}^{\infty} \log |\phi(u)|\, u^{-2} du = -\pi^2 A . \tag{3.4}$$

It follows from (3.1) that the ratio $\Lambda(t)/t$ is bounded. Hence, by (2.5) and (2.9),

$$\begin{aligned}\int_{-y}^{y} \log |\phi(u)|\, u^{-2} du &= -\frac{\pi^2}{y}\int_0^\infty N_2\left(\frac{t}{y}\right) d\Lambda(t)\\ &= \frac{\pi^2}{y}\int_0^\infty \Lambda(t) d_t N_2\left(\frac{t}{y}\right) = -2\int_0^\infty t^{-2}\Lambda(t)\log\left|\frac{y+t}{y-t}\right| dt\\ &= O\left\{\int_0^y \frac{dt}{t}\log\left|\frac{1+t/y}{1-t/y}\right| + \int_y^\infty \frac{dt}{t}\log\left|\frac{1+y/t}{1-y/t}\right|\right\} = O(1).\end{aligned}$$

Being combined with (3.2) this shows that

$$\int_{-y}^{y} \log^- |\phi(u)|\, u^{-2} du = O(1).$$

Hence the integral

$$\int_{-\infty}^{\infty} \log^- |\phi(u)|\, u^{-2} du$$

converges, whence, again by (3.2), the convergence of the integral (3.4) follows. Expression (1.2) for the constant A of Theorem I is now immediately obtained.

Another, non-Tauberian, proof of Theorem 1 proceeds as follows. We have shown in the above discussion that the integral

$$\int_{-\infty}^{\infty} |\log |\phi(u)||\, u^{-2} du$$

converges. Hence the harmonic function

$$F(u, v) = \log |\phi(u + iv)| - \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{v \log |\phi(u')|}{(u' - u)^2 + v^2} du'$$

exists in the upper half-plane $v>0$, and vanishes for $v=0$ (except at the zeros of $\phi(w)$, $u=\lambda_n^2$). It may be extended by reflection to the lower half-plane. The resulting harmonic function will be continuous everywhere, even at the zeros of $\phi(w)$. Indeed, $F(u, v)$ vanishes along the segment of the line $v=0$ through such a point, and thus cannot have a logarithmic singularity there, while the order of singularity cannot be greater than logarithmic. Thus $F(u, v)$ is the real part of an entire function.

By (1.1),

$$\log|\phi(z)| = O(|z|).$$

Now,

$$\int_{-\infty}^{\infty}\frac{v\log|\phi(u')|}{(u'-u)^2+v^2}du' = \int_{-\infty}^{-2u}+\int_{-2u}^{2u}+\int_{2u}^{\infty}$$

$$\leqq 8v\int_{-\infty}^{\infty}|\log|\phi(u')||\,u'^{-2}du' + \text{const}\int_{-2u}^{2u}\frac{u'v}{(u'-u)^2+v^2}du'$$

$$= O(|z|).$$

Thus we must have

$$F(u, v) = \pi A v,\quad v > 0,$$

for some A. Thus, as $v\to\infty$,

$$\log\phi(iv) = \log|\phi(iv)| = \pi A v + \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{v\log|\phi(u')|}{u'^2+v^2}du' = \pi A v + o(v)$$

which is the desired result.

Proof of Theorem II. If $\{z_\nu\}$ is the sequence of zeros of $f(z)$ we have

$$f(z) = \prod_{\nu=1}^{\infty}\left(1-\frac{z}{z_\nu}\right).$$

We set

$$\psi(w) = f(w^2) = \prod_{\nu=1}^{\infty}\left(1-\frac{w^2}{z_\nu}\right),$$

$$f^*(z) = \prod_{\nu=1}^{\infty}\left(1-\frac{z}{\lambda_\nu^2}\right),\quad |z_\nu| = \lambda_\nu^2,$$

$$\phi(w) = f^*(z) = \prod_{\nu=1}^{\infty}\left(1-\frac{w^2}{\lambda_\nu^2}\right),\ z = w^2.$$

Then, since $\sum_1^\infty \lambda_\nu^{-2} < \infty$,

$$n_f(r^2) = n_{f^*}(r^2) \sim \pi^{-1}\log M_{f^*}(r^2) = \pi^{-1}\log M_\phi(r) = \pi^{-1}\log\phi(ir).$$

Hence, by hypothesis (1.5) of Theorem II,

$$\log \phi(ir) \sim \pi B,$$

and, by Lemma 2.1,

$$\int_{-\infty}^{\infty} \log | \phi(u) | u^{-2} du = -\pi^2 B.$$

By hypothesis (1.4),

$$-\pi^2 B = \int_0^{\infty} \log | f(x) | x^{-3/2} dx = \int_{-\infty}^{\infty} \log | \psi(u) | u^{-2} du.$$

Thus we must have

$$\int_{-\infty}^{\infty} [\log | \psi(u) | - \log | \phi(u) |] u^{-2} du = 0. \tag{3.5}$$

On the other hand,

$$\log | \psi(u) | - \log | \phi(u) | = \sum_{\nu=1}^{\infty} \log \frac{| z_\nu - u^2 |}{| | z_\nu | - u^2 |} > 0,$$

unless all roots z_ν are real and positive. Thus relation (3.5) implies that all roots z_ν are positive, and Theorem II is proved.

4. In this paragraph we use the notation of §2, but make a slightly different assumption concerning the asymptotic behavior of $\phi(iy)$.

LEMMA 4.1. *If the series* $\sum_1^\infty \lambda_\nu^{-2}$ *converges, then the statements*

$$\log \phi(iy) \sim \pi A | y | \log | y | \ \textit{as} \ | y | \to \infty \tag{4.1}$$

and

$$\int_{-y}^{y} \log | \phi(x) | x^{-2} dx \sim -\pi^2 A \log | y | \tag{4.2}$$

are completely equivalent.

We assume $y>0$. Using the kernels $N_1(\lambda)$, $N_2(\lambda)$ of §2 we may replace (4.1), (4.2) respectively by

$$(y \log y)^{-1} \int_0^{\infty} N\left(\frac{t}{y}\right) d\Lambda(t) \to A, \ N(\lambda) = N_1(\lambda), \ N_2(\lambda). \tag{4.3}$$

We now observe that either of the statements (4.3) implies

$$\Lambda(y) = O(y \log y). \tag{4.4}$$

Indeed if (4.3) is satisfied with $N=N_1$ or $N=N_2$, then

$$\begin{aligned} O(1) &\geqq (y \log y)^{-1} \int_0^y N\left(\frac{t}{y}\right) d\Lambda(t) \\ &= N(1)(y \log y)^{-1}\Lambda(y) - (y \log y)^{-1} \int_0^y \Lambda(t) d_t N\left(\frac{t}{y}\right) \\ &> N(1)\Lambda(y)(y \log y)^{-1}, \end{aligned}$$

since $N(\lambda)$ is positive and decreasing. Next we prove that, under the condition (4.4), (4.3) is equivalent to

$$\frac{1}{y}\int_0^\infty N\left(\frac{t}{y}\right) d\Lambda^*(t) \to A, \quad N(\lambda) = N_1(\lambda),\ N_2(\lambda), \tag{4.5}$$

where

$$\Lambda^*(y) = \int_0^y (\log t)^{-1} d\Lambda(t). \tag{4.6}$$

It is readily seen from (4.4) and (4.6) that $\Lambda^*(y)$ vanishes for sufficiently small y, while

$$\Lambda^*(y) = O(y) \text{ as } y \to \infty.$$

Now the difference of the left-hand members of (4.5) and (4.3) is equal to

$$\begin{aligned} I(y) &= \frac{1}{y}\int_0^\infty N\left(\frac{t}{y}\right)\left(\frac{1}{\log t} - \frac{1}{\log y}\right) d\Lambda(t) \\ &= (y \log y)^{-1} \int_0^\infty N\left(\frac{t}{y}\right) \log \frac{y}{t} d\Lambda^*(t) \\ &= -(y \log y)^{-1} \int_0^\infty \Lambda^*(t) d_t\left[N\left(\frac{t}{y}\right) \log \frac{y}{t}\right] \\ &= O\left\{(\log y)^{-1} \int_0^\infty t \frac{d}{dt}\left[N(t) \log \frac{1}{t}\right] dt\right\} = O\left(\frac{1}{\log y}\right), \end{aligned}$$

and $\to 0$ as $y \to \infty$ or $y \to 0$. The same theorem of Wiener which was applied in the proof of Lemma 2.1 shows immediately the equivalence of the two statements (4.5); hence the two statements (4.3), and consequently (4.1) and (4.2), are also equivalent.

In order to apply Lemma 4.1 to the theory of the Riemann zeta-function we introduce

$$\Xi(z) = \xi\left(\frac{1}{2} + iz\right)$$
(4.7)
$$= \frac{1}{2}\left(\frac{1}{2} + iz\right)\left(\frac{1}{2} - iz\right)\pi^{-1/4-iz/2}\Gamma\left(\frac{1}{4} + \frac{iz}{2}\right)\zeta\left(\frac{1}{2} + iz\right).$$

It is known that $\Xi(z)$ is an entire function, is even and has all zeros in the strip $|\mathfrak{J}(z)| < \frac{1}{2}$. Moreover

$$\log \Xi(iy) = O(y) + \log \Gamma(y/2) \sim \tfrac{1}{2} y \log y,$$

$$\Xi(z) = c \prod_{\nu=1}^{\infty}\left(1 - \frac{z^2}{z_\nu^2}\right), \quad \sum_1^{\infty} |z_\nu|^{-2} < \infty, \ c = \Xi(0).$$

We set

$$z_\nu = z_\nu' + iz_\nu', \ z_\nu'' > 0, \ |z_\nu''| < \tfrac{1}{2}; \ |z_\nu| = \lambda_\nu.$$

Let us put

$$H(z) = c \prod_{\nu=1}^{\infty}\left(1 - \frac{z^2}{\lambda_\nu^2}\right).$$

We have outside the strip $|\mathfrak{J}(z)| \leqq 1$,

$$\log\left|\frac{H(z)}{\Xi(z)}\right| = -\sum_{\nu=1}^{\infty} \log\left|\frac{z_\nu^2 - z^2}{\lambda_\nu^2 - z^2}\right| = -\sum_{\lambda_\nu \leqq |2z|} - \sum_{\lambda_\nu > |2z|}$$
$$= \sum_{\lambda_\nu < |2z|} O\left(\frac{|z|}{\lambda_\nu^2}\right) + \sum_{\lambda_\nu > |2z|} O\left(\frac{|z|^2}{\lambda_\nu^2}\right) = O(|z|).$$

Thus, assuming $y > 0$,

$$\log H(iy) \sim \tfrac{1}{2} y \log y,$$

and, by Lemma 4.1,

$$(4.8) \qquad (\log y)^{-1} \int_{-y}^{y} \log |c^{-1}H(x)| \, x^{-2} dx \to -\frac{\pi}{2} \text{ as } y \to \infty.$$

Again, on the real axis,

$$\left|1 - \frac{x^2}{z_\nu^2}\right| \geqq \left|1 - \frac{x^2}{\lambda_\nu^2}\right|,$$

whence

$$\log(|1 - x^2/z_\nu^2| / |1 - x^2/\lambda_\nu^2|) \geqq 0.$$

Furthermore

$$0 \leqq I_\nu = \int_{-\infty}^{\infty} \log \frac{\left|1 - \dfrac{x^2}{z_\nu^2}\right|}{\left|1 - \dfrac{x^2}{\lambda_\nu^2}\right|} x^{-2} dx$$

$$\leqq \frac{2}{\lambda_\nu^2} \int_0^\infty \log \left| \frac{1+t^2}{1-t^2} \right| t^{-2} dt = O\left(\frac{1}{\lambda_\nu^2}\right),$$

hence we can integrate term-wise and obtain

$$0 < \int_{-\infty}^{\infty} \log \left| \frac{\Xi(x)}{\mathrm{H}(x)} \right| x^{-2} dx = \sum_{\nu=1}^{\infty} I_\nu < \infty .$$

Then, by (4.8),

$$(\log y)^{-1} \int_{-y}^{y} \log | c^{-1}\Xi(x) | \, x^{-2} dx \to -\frac{\pi}{2} \cdot$$

If we return to the zeta-function using (4.7), this gives our final result

$$\int_1^y \frac{\log | \zeta(\frac{1}{2} + ix) |}{x^2} dx = o(\log y). \tag{4.9}$$

5. Titchmarsh has† discussed asymptotic properties of entire functions with real negative zeros. In this paragraph we indicate some results which overlap with those of Titchmarsh. The method used in deriving these results is closely analogous to that used in proving Lemma 2.1; therefore we shall give here only a brief outline of the proof leaving the details to the reader.

Let

$$f(z) = \prod_{\nu=1}^{\infty} \left(1 + \frac{z}{a_\nu}\right)$$

be an entire function all of whose zeros $\{-a_\nu\}$ are negative. It will be assumed that

$$0 < a_1 \leqq a_2 \leqq \cdots, \quad \sum_{\nu=1}^{\infty} a_\nu^{-1} < \infty . \tag{5.1}$$

For simplicity we shall use the symbol $n(r)$ instead of $n_f(r)$ of the preceding paragraphs. The letter x will designate a real positive variable which tends to infinity.

† Loc. cit.

LEMMA 5.1. *Let* λ, ρ, θ *be fixed numbers such that*

$$\lambda > 0, \quad 0 < \rho < 1, \quad |\theta| < \pi. \tag{5.2}$$

Then the statements

(i) $$n(x) \sim \lambda x^\rho,$$

(ii) $$\log f(x) \sim \pi\lambda \operatorname{cosec} \pi\rho\, x^\rho,$$

(iii) $$\log |f(xe^{i\theta})| \sim \pi\lambda \operatorname{cosec} \pi\rho \cos\theta\rho\, x^\rho,$$

(iv) $$\int_0^x r^{-1-\pi/(2\theta)} \log |f(re^{i\theta})|\, dr \sim -\frac{\pi\lambda \operatorname{cosec} \pi\rho \cos\theta\rho}{\rho - \dfrac{\pi}{2\theta}} x^{\rho-\pi/(2\theta)}$$

are all equivalent. In the last statement (iv) *the right-hand member in the case* $\theta=\pi/(2\theta)$ *should be replaced by its limiting value as* $\rho\to\pi/(2\theta)$.

We first observe that the convergence of the series $\sum_1^\infty a_\nu^{-1}$ implies

$$n(x) = o(x). \tag{5.3}$$

Next let us put

$$\omega(x) = x^{-\rho} n(x). \tag{5.4}$$

In view of the fact that $n(x)$ is monotone increasing it is readily seen that the statements (i), which can be written as

$$\omega(x) \to \lambda, \tag{5.5}$$

and

$$\int_0^x \omega(r)dr \sim \lambda x, \tag{5.6}$$

are equivalent.†

Our next step is to transform the left-hand members of (ii–iv) in such a way as to allow an immediate application of Wiener's Tauberian theorems. We have

$$x^{-\rho} \log f(x) = x^{-\rho} \int_0^\infty \log\left(1 + \frac{x}{t}\right) dn(t)$$

$$= x^{-\rho} \int_0^\infty n(t) \frac{x}{t(t+x)} dt = \frac{1}{x} \int_0^\infty \omega(t) \frac{\left(\dfrac{t}{x}\right)^{\rho-1}}{1 + \dfrac{t}{x}} dt,$$

† This is readily proved directly or derived from a theorem of Wiener, loc. cit., Theorem XIII, pp. 34–35; it also follows from a well known theorem of Landau, *Beiträge zuranalytischen Zahlentheorie*, Rendiconti del Circolo Matematico di Palermo, vol. 26 (1908), pp. 169–302; p. 218.

$$x^{-\rho}\log|f(xe^{i\theta})| = \frac{1}{2}x^{-\rho}\int_0^\infty \log\left|1+\frac{x}{t}e^{i\theta}\right|^2 dn(t)$$

$$= \frac{1}{x}\int_0^\infty \omega(t)\left(\frac{t}{x}\right)^{\rho-1}\frac{1+\dfrac{t}{x}\cos\theta}{1+\dfrac{2t}{x}\cos\theta+\dfrac{t^2}{x^2}}dt,$$

$$x^{\pi/(2\theta)-\rho}\int_0^x r^{-1-\pi/(2\theta)}\log|f(re^{i\theta})|\,dr$$

$$= x^{\pi/(2\theta)-\rho}\int_0^x r^{-1-\pi/(2\theta)}dr\, r^{\rho-1}\int_0^\infty \omega(t)\left(\frac{t}{r}\right)^{\rho-1}\frac{1+\dfrac{t}{r}\cos\theta}{1+\dfrac{2t}{r}\cos\theta+\dfrac{t^2}{r^2}}dt$$

$$= -\frac{1}{x}\int_0^\infty \omega(t)dt\left(\frac{x}{t}\right)^{1-\rho+\pi/(2\theta)}\int_{x/t}^\infty \frac{1+\dfrac{1}{r}\cos\theta}{1+\dfrac{2}{r}\cos\theta+\dfrac{1}{r^2}}r^{-1-\pi/(2\theta)}dr$$

(see (5.9) below, for $u=0$, $\theta\rho=\pi/2$). Thus all the statements (ii–iv) are expressible in the form

$$\frac{1}{x}\int_0^\infty \omega(t)N\left(\frac{t}{x}\right)dt \to \lambda \text{ as } x\to\infty\ , \tag{5.7}$$

where $N(y)$ stands, respectively, for

$$N_3(y) = \frac{1}{\pi\operatorname{cosec}\pi\rho}\,\frac{y^{\rho-1}}{1+y},$$

$$N_4(y) = \frac{1}{\pi\operatorname{cosec}\pi\rho\cos\theta\rho}\,y^{\rho-1}\frac{1+y\cos\theta}{1+2y\cos\theta+y^2},$$

$$N_5(y) = -\frac{\rho-\dfrac{\pi}{2\theta}}{\pi\operatorname{cosec}\pi\rho\cos\theta\rho}\,y^{\rho-1-\pi/(2\theta)}\int_{1/y}^\infty \frac{1+\dfrac{1}{r}\cos\theta}{1+\dfrac{2}{r}\cos\theta+\dfrac{1}{r^2}}r^{-1-\pi/(2\theta)}dr.$$

A direct computation yields

$$\int_0^\infty \frac{y^{iu+\rho-1}}{1+y}dy = \pi \operatorname{cosec} \pi(iu+\rho), \tag{5.8}$$

$$\begin{aligned} &\int_0^\infty \frac{y^{iu+\rho-1}(1+y\cos\theta)}{1+2y\cos\theta+y^2}dy \\ &= \frac{1}{2}\left[\int_0^\infty \frac{y^{iu+\rho-1}}{1+ye^{i\theta}}dy + \int_0^\infty \frac{y^{iu+\rho-1}}{1+ye^{-i\theta}}dy\right] \\ &= \int_0^\infty \frac{y^{iu+\rho-1}}{1+y}\,\frac{1}{2}[(e^{i\theta})^{-iu-\rho}+(e^{-i\theta})^{-iu-\rho}]dy \\ &= \pi \operatorname{cosec} \pi(iu+\rho)\cos\theta(iu+\rho), \end{aligned} \tag{5.9}$$

$$\begin{aligned} &\int_0^\infty y^{iu+\rho-1-\pi/(2\theta)}dy\int_{1/y}^\infty r^{-1-\pi/(2\theta)}\frac{1+\dfrac{1}{r}\cos\theta}{1+\dfrac{2}{r}\cos\theta+\dfrac{1}{r^2}}dr \\ &= \int_0^\infty r^{-1+\pi/(2\theta)}\frac{1+r\cos\theta}{1+2r\cos\theta+r^2}dr\int_r^\infty y^{iu+\rho-1-\pi/(2\theta)}dy \\ &= -\frac{\pi\operatorname{cosec}\pi(iu+\rho)\cos\theta(iu+\rho)}{iu+\rho-\dfrac{\pi}{2\theta}}. \end{aligned} \tag{5.10}$$

The last result is first derived in the case where $0<\rho<1+\pi/(2\theta)$, but is easily extended to the general case $0<\rho<1$ by analytic continuation. It is an easy matter to verify that the kernels $N_3(y)$, $N_4(y)$ when $|\theta|<\pi/2$, and $N_5(y)$ are possessed of all the properties of the kernel $N(y)$ stated in the proof of Lemma 2.1. We set $\Lambda(t)=\int_0^t\omega(t)dt$. Since $\omega(t)\geqq 0$, $\Lambda(t)$ is monotone increasing. Hence Wiener's theorem used in §2 may be applied here with the result that the statements (ii), (iii) when $|\theta|<\pi/2$, and (iv) are equivalent, while either of (ii) or (iv) implies (iii) when $\pi/2<|\theta|<\pi$. It should be observed that the kernel $N_4(y)$ is not positive when $|\theta|>\pi/2$, while $N_5(y)$ is positive over the whole range $|\theta|<\pi$. The introduction of this kernel was necessitated by the lack of positiveness of $N_4(y)$ when $|\theta|>\pi/2$. Another theorem of Wiener† will show that either of the statements (ii), (iii) when $|\theta|<\pi/2$, and (iv) implies (5.6), hence (5.5) which is the same as (i). On the other hand it may be proved directly‡ that (i) implies (ii), hence also (iii) and (iv). This completes the proof of Lemma 5.1.

† N. Wiener, loc. cit., Theorem XI″, pp. 31–32.

‡ Titchmarsh, loc. cit., Theorem I.

VI. On two problems of Pólya

1. Pólya† has set the following problem: *Let the real numbers* $m_1, m_2, \cdots$ *have the properties* $0<m_1<m_2<\cdots$ *and*

$$\underline{\lim}_{n\to\infty} \frac{n}{m_n} > \frac{b-a}{2\pi} > 0. \tag{1.1}$$

Furthermore, let $f(x)$ *be continuous in the closed interval* $[a, b]$. *Then it will follow from*

$$\int_a^b f(x)\cos m_n x\,dx = \int_a^b f(x)\sin m_n x\,dx = 0 \tag{1.2}$$

that $f(x)$ *vanishes identically.*

There is no restriction in supposing $b=-a=\pi$. We shall prove the following more general theorem:

Theorem I. *Let* $0<m_1<m_2<\cdots$ *and let*

$$\overline{\lim}_{n\to\infty} \frac{n}{m_n} > 1. \tag{1.3}$$

Then if $f(x)$ *belongs to* L^2 *and*

$$\int_{-\pi}^{\pi} f(x)e^{\pm i m_n x}dx = 0 \qquad (n=1, 2, 3, \cdots), \tag{1.4}$$

$f(x)$ *vanishes except over a set of measure zero.*

It is very important that we have replaced $\underline{\lim}$ by $\overline{\lim}$. This yields us a much deeper theorem.

Since (1.4) is satisfied with $f(x)$ replaced by $f(x)\pm f(-x)$, it suffices to consider only the cases of $f(x)$ even or odd. We shall give the discussion of the case $f(x)$ even, under the additional assumption that $\int_{-\pi}^{\pi} f(t)dt\neq 0$. The case where this assumption is not satisfied as well as the case of $f(x)$ odd will require but slight modifications which may be left to the reader. We set

$$\phi(u) = \int_{-\pi}^{\pi} f(t)e^{iut}dt, \tag{1.5}$$

where the entire function $\phi(u)$ is even and where we may assume without loss of generality that $\phi(0)=1$. We observe that, on setting $u=\sigma+i\tau$, we have

† Jahresbericht der Deutschen Mathematiker-Vereinigung, vol. 40 (1931), Abteilung 2, p. 81, Problem 108.

$$(1.6)\qquad |\phi(u)| = |\phi(\sigma + i\tau)| \leqq \left\{\int_{-\pi}^{\pi} |f(t)|^2 dt\right\}^{1/2} e^{\pi|\tau|} = O(e^{\pi|\tau|}).$$

On the other hand by the theory of Fourier transforms we know that $\phi(\sigma) \subset L^2$ over $(-\infty, \infty)$ and that the Fourier transform of $\phi(\sigma)$ vanishes outside $(-\pi, \pi)$. Hence by Theorem II of our Note I,†

$$(1.7)\qquad \int_{-\infty}^{\infty} \frac{|\log|\phi(\sigma)||}{1+\sigma^2} d\sigma < \infty.$$

By the change of variable $u^2 = z$ we obtain a function $\psi(z) = \phi(u)$ which satisfies the conditions of Theorem I of our preceding Note V. It follows at once that the limit

$$(1.8)\qquad \lim_{r\to\infty} \frac{n_\phi(r)}{r} = A$$

exists. Let $\{u_\nu\}$ be the sequence of zeros of $\phi(u)$. It is clear that $\{\pm m_\nu\}$ is a subsequence of $\{u_\nu\}$. Hence, by (1.3),

$$(1.9)\qquad 2 < \overline{\lim_{n\to\infty}} \frac{2n}{m_n} \leqq \lim_{r\to\infty} \frac{n_\phi(r)}{r} = A.$$

However, by Jensen's theorem, in view of (1.6),

$$(1.10)\qquad \begin{aligned} \frac{1}{r}\int_0^r \frac{n_\phi(t)}{t} dt &= \frac{1}{2\pi r}\int_0^{2\pi} \log|\phi(re^{i\theta})|\, d\theta \\ &\leqq \frac{1}{2\pi r}\int_0^{2\pi} \pi r|\sin\theta|\, d\theta + O\left(\frac{1}{r}\right) = 2 + O\left(\frac{1}{r}\right). \end{aligned}$$

Hence

$$(1.11)\qquad A = \lim_{r\to\infty} \frac{1}{r}\int_0^r \frac{n_\phi(t)}{t} dt \leqq 2.$$

The resulting contradiction shows that $f(x)$ must vanish except for a set of measure zero.

2. Pólya‡ has also set the following problem: *Let $f(z)$ be an entire function which is bounded for the integral arguments $z = 0, \pm 1, \pm 2, \cdots, \pm n, \cdots$. Let*

$$(2.1)\qquad M_f(r) = o(r).$$

Then $f(z)$ reduces to a constant.

† The present volume of these Transactions, p. 349.

‡ Jahresbericht der Deutschen Mathematiker-Vereinigung, vol. 40 (1931), Abteilung 2, p. 80, Problem 105.

It is clearly sufficient to prove the theorem for an even $f(z)$, for if $f(z)$ is odd, we need only consider $f(z)/z$, which will be even, and will hence reduce to a constant which can only be zero. The general function may then be treated by reducing it to the sum of an odd and an even part.

If $f(z)$ is even,

$$g(z) = [f(z) - f(0)]z^{-2} \tag{2.2}$$

will be entire. Thus

$$\sum_{n=-\infty}^{\infty} |\, g(n)\,| < \infty\ . \tag{2.3}$$

Let us form

$$G(z) = \sum_{n=-\infty}^{\infty} g(n)\,\frac{\sin \pi(n-z)}{\pi(n-z)}\,. \tag{2.4}$$

Clearly

$$G(x+iy) = O(y^{-1}e^{\pi|y|}). \tag{2.5}$$

Let us now form the entire function

$$H(z) = [g(z) - G(z)] \operatorname{cosec} \pi z. \tag{2.6}$$

For all values of z and all integral values of n we shall have

$$\begin{aligned} H[(n+\tfrac{1}{2}) + iy] &= O\{\exp(\epsilon[(n+\tfrac{1}{2})^2 + y^2]^{1/2} - \pi|\,y\,|)\} \\ &\quad + O(1/|y|) \\ &= e^{\epsilon|n+1/2|}O(e^{(\epsilon-\pi)|y|}) + O(1/|\,y|) \end{aligned} \tag{2.7}$$

uniformly in n. We have here employed (2.1) and (2.5). Hence

$$\int_{-\infty}^{\infty} |\, H(n+\tfrac{1}{2}+iy)\,|^2 dy = O(e^{2\epsilon|n|}), \tag{2.8}$$

for all ϵ.

Let us put

$$x_1 = [x+\tfrac{1}{2}],\ x_2 = [x-\tfrac{1}{2}].$$

Then, by Cauchy's theorem,

$$\begin{aligned} H(x+iy) &= (2\pi i)^{-1}\int_{-\infty}^{\infty} \frac{H(x_1+\frac{1}{2}+iy_1)}{x_1+\frac{1}{2}+iy_1-x-iy}\,dy_1 \\ &\quad - (2\pi i)^{-1}\int_{-\infty}^{\infty} \frac{H(x_2-\frac{1}{2}+iy_1)}{x_2-\frac{1}{2}+iy_1-x-iy}\,dy_1. \end{aligned} \tag{2.9}$$

Hence

$$(2.10)\quad \int_{-\infty}^{\infty} H(x+iy)e^{iuy}dy$$

$$= (2\pi i)^{-1}\int_{-\infty}^{\infty} H\left(x_1+\frac{1}{2}+iy_1\right)e^{iuy_1}dy_1\int_{-\infty}^{\infty}\frac{e^{iuy}dy}{x_1+\frac{1}{2}-x-iy}$$

$$-(2\pi i)^{-1}\int_{-\infty}^{\infty} H\left(x_2-\frac{1}{2}+iy_1\right)e^{iuy_1}dy_1\int_{-\infty}^{\infty}\frac{e^{iuy}}{x_2-\frac{1}{2}-x-iy},$$

and, by the Plancherel theorem and the Schwarz unequality,

$$\int_{-\infty}^{\infty}|H(x+iy)|^2dy \leqq \text{const}\left\{\int_{-\infty}^{\infty}|H(x_1+\tfrac{1}{2}+iy)|^2dy\right.$$
$$\left.+\int_{-\infty}^{\infty}|H(x_2-\tfrac{1}{2}+iy)|^2dy\right\}.$$

Thus, by (2.8),

$$(2.11)\qquad \int_{-\infty}^{\infty}|H(x+iy)|^2dy = O(e^{2\epsilon|x|}).$$

By an application of Cauchy's theorem,

$$(2.12)\qquad \int_{-\infty}^{\infty}H(x+iy)e^{iuy}dy = e^{-ux}\int_{-\infty}^{\infty}H(iy)e^{iuy}dy \equiv e^{-ux}\phi(y).$$

Thus by the Plancherel theorem,

$$(2.13)\qquad \int_{-\infty}^{\infty}|\phi(u)|^2e^{-2ux}du = O(e^{2\epsilon|x|}).$$

This is however only possible if $\phi(u)$ vanishes almost everywhere for $|u|>\epsilon$. Since ϵ is arbitrarily small, $\phi(u)$ must be equivalent to zero. Thus $H(z)$ vanishes, and $g(z)=G(z)$. On the other hand,

$$(2.14)\qquad M_G(r)\sim \pi r,$$

unless every $g(n)$ is zero. This yields

$$(2.15)\qquad G(z) = g(z) = 0,\ f(z) = f(0),$$

which is the desired result.

VII. On the Volterra equation

1. A theorem of Mercer† asserts that *if* $0<\alpha<1$, *and if*

† J. Mercer, *On the limits of real variants*, Proceedings of the London Mathematical Society, (2), vol. 5 (1907), pp. 206–224.

$$\alpha s_n + (1-\alpha)\frac{1}{n}\sum_{1}^{n} s_\nu \to s, \tag{1.1}$$

then

$$s_n \to s. \tag{1.2}$$

This theorem possesses generalizations of a non-trivial nature. The continuous analogue asserts that if $0<\alpha<1$ and if

$$\alpha s(x) + \frac{1-\alpha}{x}\int_1^x s(y)dy \to s \text{ as } x\to\infty, \tag{1.3}$$

then

$$s(x) \to s. \tag{1.4}$$

By a change of independent variable, this asserts that if

$$\alpha S(\xi) + (1-\alpha)\int_0^\xi e^{\eta-\xi}S(\eta)d\eta \to s, \tag{1.5}$$

then

$$S(\xi)\to s. \tag{1.6}$$

This statement is a particular case of the following theorem:

THEOREM I. *Let $F(x)$ be measurable and bounded over every finite range* $(0, A)$. *Let $K(x)\subset L$ over* $(0,\infty)$, *that is,*

$$\int_0^\infty |K(\xi)|\,d\xi < \infty\ . \tag{1.7}$$

Let

$$F(x) + \int_0^x K(x-\xi)F(\xi)d\xi \to s \textit{ as } x\to\infty\ . \tag{1.8}$$

Then if

$$\int_0^\infty K(\xi)e^{-w\xi}d\xi \neq -1,\ \Re(w)\geqq 0, \tag{1.9}$$

we shall have

$$F(x)\to s\left[1+\int_0^\infty K(\xi)d\xi\right]. \tag{1.10}$$

Conversely, let $K(x)\subset L$, $\int_0^\infty K(\xi)d\xi\neq -1$, and let (1.8) *imply* (1.10) *for every $F(x)$ satisfying our conditions. Then* (1.9) *must be true.*

2. The second part of this theorem may be proved by reductio ad absurdum by putting

$$F(x) = e^{w_0 x},$$

where

$$\int_0^\infty K(\xi)e^{-w_0\xi}d\xi = -1,\ \Re(w_0) \geqq 0,\ w_0 \neq 0.$$

Then

$$\left|F(x) + \int_0^x K(x-\xi)F(\xi)d\xi\right| = \left|e^{w_0 x}\int_x^\infty K(\xi)e^{-w_0\xi}d\xi\right|$$

$$\leqq \int_x^\infty |K(\xi)|\,d\xi \to 0.$$

As (1.10) is obviously false, the second part of the theorem is proved.

3. The first part of Theorem I will appear as a corollary to a theorem concerning the Volterra integral equation of the closed cycle.

We shall use the symbol

$$A \star B(x)$$

to indicate the "Faltung" of the two functions $A(x)$, $B(x)$,

$$A \star B(x) = \int_0^x A(\xi)B(x-\xi)d\xi = \int_0^x A(x-\xi)B(\xi)d\xi = B \star A(x).$$

It is well known that the (bounded and measurable) solution of the Volterra integral equation

$$G(x) = F(x) + K \star F(x) \tag{3.1}$$

is uniquely determined and given by

$$F(x) = G(x) + Q \star G(x) \tag{3.2}$$

where the resolvent kernel $Q(x)$ itself is determined from

$$Q(x) + K(x) = -K \star Q(x) = -Q \star K(x), \tag{3.3}$$

or else, by

$$Q(x) = \sum_{n=1}^{\infty}(-1)^n K^n(x),\ K^n(x) = K \star K \star \cdots \star K(x). \tag{3.4}$$

We observe that the solution of (3.3) is easily obtained by using Laplace transforms.† Let us designate by

† See, e.g., S. Bochner, *Vorlesungen über Fouriersche Integrale*, Leipzig, 1932, chapter VII. Other references are also found there.

$$k(w) = \int_0^\infty K(\xi)e^{-w\xi}d\xi, \tag{3.5}$$

$$q(w) = \int_0^\infty Q(\xi)e^{-w\xi}d\xi, \tag{3.6}$$

the Laplace transforms of $K(x)$, $Q(x)$. Equation (3.4) reduces then to

$$q(w) = \frac{-\,k(w)}{1 + k(w)}, \tag{3.7}$$

and $Q(x)$ will be found by inversion of a Laplace integral.

The theorem in question may now be stated as follows:

THEOREM II. *A necessary and sufficient condition that* $Q(x) \subset L$ *over* $(0, \infty)$, *that is,*

$$\int_0^\infty |\,Q(\xi)\,|\,d\xi < \infty\,, \tag{3.8}$$

is that

$$k(w) = \int_0^\infty K(\xi)e^{-w\xi}d\xi \neq -\,1, \quad \Re(w) \geqq 0. \tag{3.9}$$

If this theorem holds true, the first part of Theorem I is immediately derived. Indeed, under the assumptions made we have

$$F(x) = G(x) + \int_0^x G(x - \xi)Q(\xi)d\xi.$$

Here $G(\xi)$ is bounded over every finite range and $\rightarrow s$ as $\xi \rightarrow \infty$. Hence $G(\xi)$ is bounded over the whole range $(0, \infty)$. Since $Q(\xi)$ is integrable over $(0, \infty)$ we may pass to limit as $x \rightarrow \infty$ under the integral sign, with the result

$$F(x) \rightarrow s + s\int_0^\infty Q(\xi)d\xi = s[1 + q(0)] = s[1 + k(0)]^{-1}$$

which is precisely the desired formula (1.10).

4. To prove the necessity of (3.9) we observe that if (3.8) holds then $q(w)$ as well as $k(w)$ are analytic in the half-plane $\Re(w) > 0$ and continuous up to and including the boundary $\Re(w) = 0$. This implies that the denominator in the right-hand member of (3.7) does not vanish for $\Re(w) \geqq 0$, so that (3.9) holds.

The proof that (3.9) is sufficient is more difficult. We introduce the auxiliary functions

$$\phi_A(u) = \begin{cases} 1, \ |u| < A, \\ 2 - |u|/A, \ A \leqq |u| \leqq 2A, \\ 0, \ |u| > 2A, \end{cases} \tag{4.1}$$

and put

$$q^*(w) \equiv \frac{-k(w)}{1 + k(w)}, \tag{4.2}$$

$$q^*(iu) = q_1(u) + q_2(u), \tag{4.3}$$

$$q_1(u) = \phi_A(u)q^*(iu), \ q_2(u) = [1 - \phi_A(u)]q^*(u). \tag{4.4}$$

We wish to prove that if A is sufficiently large, $q_1(u)$ and $q_2(u)$ are both Fourier transforms of functions of L.

To begin with,

$$q_1(u) = \begin{cases} \dfrac{-\phi_A(u)k(iu)}{\phi_{2A}(u) + \phi_{2A}(u)k(iu)} \text{ when } |u| < 2A, \\ 0 \text{ when } |u| \geqq 2A. \end{cases}$$

Thus $q_1(u)$ is the quotient of two functions, each the Fourier transform of a function of L, each vanishing outside a finite range, and such that the denominator function only vanishes in points interior to the region of vanishing of the numerator function. We may then appeal to a theory due to Wiener† to show that $q_1(u)$ is the Fourier transform of a function of L.

We have

$$q_2(u) = [1 - \phi_A(u)]\frac{-k(iu)[1 - \phi_{A/2}(u)]}{1 + k(iu)[1 - \phi_{A/2}(u)]}.$$

It is easy to show that this is the Fourier transform of a function of L when the same is true of

$$-k(iu)[1 - \phi_{A/2}(u)]\{1 + k(iu)[1 - \phi_{A/2}(u)]\}^{-1}$$
$$= \sum_{n=1}^{\infty}(-1)^n\{k(iu)[1 - \phi_{A/2}(u)]\}^n.$$

Now

$$\{k(iu)[1 - \phi_{A/2}(u)]\}^n$$

is the Fourier transform of a function $h_n(x)$ for which

† N. Wiener, *The Fourier Integral and Certain of its Applications*, Cambridge, 1933; Lemmas 6_7, 6_{10}, 6_{18}.

$$\int_{-\infty}^{\infty} | h_n(\xi) | d\xi \leqq \left[\int_{-\infty}^{\infty} | h_1(\xi) | d\xi \right]^n$$

$$= \left[\int_{-\infty}^{\infty} d\xi \left| K(\xi) - \frac{1}{\pi A} \int_0^{\infty} K(\eta) \frac{\cos \frac{A}{2}(\xi - \eta) - \cos A(\xi - \eta)}{(\xi - \eta)^2} d\eta \right| \right]^n.$$

An argument of the familiar Fejér type will show that we may choose A so large that the integral in brackets is less than any given number λ, $0<\lambda<1$. It will follow at once that $q_2(u)$ is the Fourier transform of a function $F_2(x)$ for which

$$\int_{-\infty}^{\infty} | F_2(\xi) | d\xi \leqq \frac{\lambda}{1 - \lambda} \cdot$$

Combining this with the similar result for $q_1(u)$, we see that we may write

$$q^*(iu) = \frac{- k(iu)}{1 + k(iu)} = \int_{-\infty}^{\infty} F(\xi) e^{-iu\xi} d\xi, \quad F(\xi) \subset L. \tag{4.5}$$

We may rewrite (4.5) as

$$\int_{-\infty}^{0} F(\xi) e^{-iu\xi} d\xi = - \int_{0}^{\infty} F(\xi) e^{-iu\xi} d\xi + q^*(iu). \tag{4.6}$$

Now, it is readily seen that $k(w) \to 0$ as $|w| \to \infty$, uniformly in the half-plane $\Re(w) \geqq 0$. Since, by hypothesis, $1 + k(w) \neq 0$ for $\Re(w) \geqq 0$, there exists a positive constant c such that

$$| 1 + k(w) | \geqq c > 0.$$

Thus

$$- \int_{0}^{\infty} F(\xi) e^{-w\xi} d\xi + q^*(w)$$

is a function of w analytic and bounded in the right half-plane, and continuous up to and including the imaginary axis. Similarly,

$$\int_{-\infty}^{0} F(\xi) e^{-w\xi} d\xi$$

is a function of w analytic and bounded in the left half-plane, and continuous up to and including the imaginary axis. Furthermore, the two functions are identical on the imaginary axis. By the classical argument of Riemann-Painlevé it readily results that they are parts of the same analytic function, which is thus entire and bounded. It hence reduces to a constant, and since

$$\int_{-\infty}^{0} F(\xi)e^{-w\xi}d\xi \to 0 \text{ as } w \to -\infty, \tag{4.7}$$

this constant can only be 0. Thus

$$q^*(w) = \frac{-k(w)}{1+k(w)} = \int_0^{\infty} F(\xi)e^{-w\xi}d\xi. \tag{4.8}$$

On the other hand, it follows readily from (3.4) that there exists a $w_0 > 0$ such that

$$\frac{-k(w)}{1+k(w)} = \int_0^{\infty} Q(\xi)e^{-w\xi}d\xi, \ \Re(w) > w_0, \tag{4.9}$$

and

$$\int_0^{\infty} |Q(\xi)| e^{-w_0\xi}d\xi < \infty . \tag{4.10}$$

By the uniqueness theorem for Laplace transforms we conclude that $F(x)e^{-wx}$ and $Q(x)e^{-wx}$ coincide almost everywhere, whence $Q(x) \subset L$.

5. In this proof, we have used the theorem† of Wiener that if a function has an absolutely convergent Fourier series and does not vanish, its reciprocal has an absolutely convergent Fourier series. P. Lévy‡ has pointed out that the same methods suffice for the following theorem: if a function $f(x)$ has an absolutely convergent Fourier series, and $\Phi(u)$ is analytic over the range of values of $f(x)$, then $\Phi[f(x)]$ has an absolutely convergent Fourier series. By methods not essentially different from those of this paper, we may extend this theorem as follows: if $f(x)$ is the Fourier transform of a function of L, and $\Phi(u)$ is analytic over the range of values of $f(x)$, including 0, then

$$\Phi[f(x)]$$

is the Fourier transform of a function of L.

---

† Loc. cit., Lemma 6_{16}.

‡ P. Lévy, *Sur la convergence absolue des séries de Fourier*, Paris Comptes Rendus, vol. 196 (1933), p. 463.

Massachusetts Institute of Technology,
Cambridge, Mass.

INFINITE SYSTEMS OF ORDINARY DIFFERENTIAL EQUATIONS WITH APPLICATIONS TO CERTAIN SECOND-ORDER PARTIAL DIFFERENTIAL EQUATIONS*

BY

DANIEL C. LEWIS, JR.

INTRODUCTION

From a purely formal point of view, the problem of integrating the non-linear partial differential equation

$$\frac{\partial^2 u}{\partial t^2} - \frac{\partial^2 u}{\partial y^2} = F\left(\frac{\partial u}{\partial y}, \frac{\partial u}{\partial t}, u, y, t\right),$$

under the conditions $u(0, t) = u(\pi, t) = 0$, $u(y, 0) = f(y)$, $u_t(y, 0) = g(y)$ (where f and g are prescribed functions) can be reduced in the following way to the problem of integrating an infinite system of ordinary differential equations,

$$\frac{d^2 x_n}{dt^2} + n^2 x_n = f_n\left[t, x_1, \frac{dx_1}{dt}, x_2, \frac{dx_2}{dt}, \cdots\right] \qquad (n = 1, 2, 3, \cdots).$$

We want the solution to be valid in a rectangular region, $0 \leqq y \leqq \pi$, $0 \leqq t \leqq K > 0$. We assume the trigonometric developments

$$f(y) = \sum_{k=1}^{\infty} a_k \sin ky, \quad g(y) = \sum_{k=1}^{\infty} a_k' \sin ky, \quad u(y, t) = \sum_{k=1}^{\infty} x_k(t) \sin ky,$$

where the a_k and a_k' are known constants and the $x_k(t)$ are unknown functions. If we formally differentiate the series for $u(y,t)$, substitute in the partial differential equation, multiply through by $(2/\pi) \sin ny$, and integrate with respect to y from 0 to π, making use of the orthogonal properties of the sine functions, we get the nth equation of the infinite system written above with

$$f_n\left[t, x_1, \frac{dx_1}{dt}, x_2, \frac{dx_2}{dt}, \cdots\right]$$
$$= \frac{2}{\pi} \int_0^{\pi} F\left[\sum_{k=1}^{\infty} x_k k \cos ky, \sum_{k=1}^{\infty} \frac{dx_k}{dt} \sin ky, \sum_{k=1}^{\infty} x_k \sin ky, y, t\right] \sin ny\, dy.$$

* Presented to the Society, December 27, 1932; received by the editors September 20, 1932, and in revised form and with addition of Part III, Feburary 7, 1933.

We must evidently solve our infinite system under the initial conditions $x_k(0)=a_k,\ dx_k/dt|_{t=0}=a_k'$.

It is the principal object of this paper to put the above formal procedure upon a rigorous basis.

In Part I, we shall study the slightly more general system

$$\frac{d^2x_n}{dt^2}+\mu_n^2x_n=f_n\left[t,\ x_1,\ \frac{dx_1}{dt},\ x_2,\ \frac{dx_2}{dt},\ \cdots\right]\qquad(n=1,2,\cdots),$$

the μ_n being arbitrary positive constants, together with the initial conditions given above. Actually we shall study this system in the equivalent integral form,

$$x_n(t)=a_n\cos\mu_nt+(a_n'/\mu_n)\sin\mu_nt$$
$$+(1/\mu_n)\int_0^t f_n\left[\tau,\ x_1(\tau),\ \frac{dx_1}{dt}(\tau),\ \cdots\right]\sin\mu_n(t-\tau)d\tau.$$

In Part II, we shall apply the results of Part I to partial differential equations, thus obtaining an existence theorem.

This plan has already been carried out by L. Lichtenstein* for equations of considerably more restricted type. The right hand side of Lichtenstein's equation is, in fact, independent of $\partial u/\partial y$ and $\partial u/\partial t$ and can be developed in a power series in u:

$$F\left(\frac{\partial u}{\partial y},\ \frac{\partial u}{\partial t},\ u,\ y,\ t\right)=\sum_{k=1}^{\infty}p_k(y,t)u^k.$$

On the other hand, the essential requirement laid down by us is that F should obey a certain Lipschitz condition in its first three arguments. The present results also represent a generalization beyond Lichtenstein's work in that the requirements on the initial values, $f(y)$ and $g(y)$, are much less restrictive. Here it is merely assumed that $f'(y)$ and $g(y)$ have summable squares on $0\leqq y\leqq\pi$, or in other words that $\sum k^2a_k^2$ and $\sum a_k'^2$ converge; whereas Lichtenstein assumes the convergence of $\sum k^2|a_k|$ and $\sum k|a_k'|$. The generalizations that Lichtenstein does carry through in other directions (as to the shape of the region and the nature of the end or "boundary" conditions) can equally well be carried out here.

On the other hand our generalizations are gained at a certain sacrifice. The solution $u(y,t)$ produced by Lichtenstein is a solution in the ordinary sense, whereas the $u(y,t)$ produced by us may be a solution only in a certain generalized sense to be defined later. This generalized notion of a solution of a partial differential equation is, however, a natural one, and has been used

* See bibliography at the end of this introduction.

by other authors. N. Wiener,* for example, has given a generalization, which, while not assuming the existence of the first derivatives, $\partial u/\partial y$ and $\partial u/\partial t$, applies only to linear equations. He gives references to Bôcher and G. C. Evans. My own definition requires the existence of the first derivatives, but, so far as I know, it is the only one which applies to the general second-order partial differential equation, linear or not.

A bibliography of the literature on infinite systems of differential equations appears at the end of this introduction. This bibliography is complete so far as I have been able to ascertain. None of the work there listed, with the exception of Lichtenstein's and Siddiqi's, can be applied here. The reason is that the usual existence theorems for infinite systems of differential equations of the form $dz_k/dt=\zeta_k(t, z_1, z_2, \cdots)$ with initial conditions $z_k(0)=c_k$, assume a too restrictive correspondence between the laws of decrease of the $|z_k-c_k|$ and the $|\zeta_k|$. This correspondence is roughly of the nature that the convergence of $\sum_k|z_k-c_k|^2$ implies the convergence of $\sum_k|\zeta_k|^2$ for t suitably restricted. Evidently such an assumption fails to take into account even the following highly degenerate example which can be integrated immediately:

$$\frac{dz_{2n-1}}{dt} = nz_{2n} \equiv \zeta_{2n-1}(t, z_1, z_2, \cdots), \qquad z_k(0) = c_k,$$

$$\frac{dz_{2n}}{dt} = -nz_{2n-1} \equiv \zeta_{2n}(t, z_1, z_2, \cdots) \quad (n, k = 1, 2, 3, \cdots).$$

Assume the convergence of $\sum_k c_k^2$. Then the convergence of $\sum_k z_k^2$ would ensure the convergence of $\sum_k|z_k-c_k|^2$, but not that of $\sum_k\zeta_k^2$. Nevertheless such an infinite system is extremely useful in the applications to partial differential equations. This particular simple system is included in the theories presented both by Lichtenstein and by me. For it may be written in the form

$$\frac{d^2x_n}{dt^2} + n^2x_n = 0;$$

if we set $x_n=z_{2n-1}$,

$$\frac{dx_n}{dt} = nz_{2n}.$$

But it can be easily shown that a large field still awaits exploration.

The infinite systems considered in this paper are formally quite like those treated by Lichtenstein and quite unlike those treated by W. L. Hart in his paper of 1922. Nevertheless the methods are much more similar to Hart's

* Mathematische Annalen, vol. 95 (1926), p. 582.

methods than to Lichtenstein's; and the author wishes to acknowledge here his less obvious debt to Hart.

The application of the results of Part I are probably not limited to the problems considered in Part II. Instead of using the trigonometric expansions, exclusively considered in Part II, one might use general Sturm-Liouville orthogonal functions. Such a procedure might furnish theories for non-linear normal hyperbolic equations (in any number of independent variables) with boundary conditions of a much more complicated type than those considered here. Here also is a large field awaiting exploration.

Existence theorems for the Cauchy problem with non-analytic initial conditions have not yet been given for general non-linear* hyperbolic equations, except for the case of two independent variables, which has been most elegantly treated by H. Lewy.† It may be that the method of infinite systems of ordinary differential equations will furnish the key to the problem. Even in the case of two independent variables Lewy's work is applicable only to the unmixed Cauchy problem, whereas this method is applicable to the mixed problem, where boundary conditions as well as initial conditions play a prominent rôle. Further developments await more general existence theorems for infinite systems of differential equations.

Bibliography of the theory of infinite systems of differential equations

H. von Koch, *Sur les systèmes d'ordre infini d'équations différentielles*, Öfversigt af Kongliga Vetenskaps-Akademiens Förhandlingar, vol. 56 (1899), pp. 395–411 (analytic non-linear theory).

E. H. Moore, *New Haven Mathematical Colloquium*, 1906 (a linear theory in the sense of "general analysis").

F. R. Moulton, *Solution of an infinite system of differential equations of the analytic type*, Proceedings of the National Academy of Sciences, vol. 1 (1915), pp. 350–354 (analytic non-linear theory). The same work is published in the text-book on differential equations by the same author.

T. H. Hildebrandt, *On a theory of linear differential equations in general analysis*, these Transactions, vol. 18 (1917), pp. 73–96.

W. L. Hart, *Differential equations and implicit functions in infinitely many*

* Riemann and Hadamard have laid the foundation for the linear case. See the latter's book *Lectures on Cauchy's Problem*.

See also M. Mathisson, *Eine neue Lösungsmethode für Differentialgleichungen von normalen hyperbolischen Typus*, Mathematische Annalen, vol. 107 (1932), pp. 400–419.

† *Über das Anfangswertproblem einer hyperbolischen nichtlinearen partiellen Differentialgleichung zweiter Ordnung mit zwei unabhängigen Veränderlichen*, Mathematische Annalen, vol. 98 (1927), pp. 179–191.

variables, these Transactions, vol. 18 (1917), pp. 125–160; *Functions of infinitely many variables in Hilbert space*, these Transactions, vol. 23 (1922), pp. 30–50 (non-analytic non-linear theory); *Linear differential equations in infinitely many variables*, American Journal of Mathematics, vol. 39 (1917), pp. 407–424; *The Cauchy-Lipschitz method for infinite systems of differential equations*, American Journal of Mathematics, vol. 43 (1921), pp. 226–231.

I. A. Barnett, *Differential equations with a continuous infinitude of variables*, American Journal of Mathematics, vol. 44 (1922), pp. 172–190; *Linear partial differential equations with a continuous infinitude of variables*, American Journal of Mathematics, vol. 45 (1923), pp. 42–53.

A. Wintner, *Zur Theorie der unendlichen Differentialsysteme*, Mathematische Annalen, vol. 95 (1925), pp. 544–556; *Zur Lösung von Differentialsystemen mit unendlich vielen Veränderlichen*, Mathematische Annalen, vol. 98 (1927), pp. 273–280 (analytic non-linear theory); *Zur Analysis im Hilbertschen Raume*, Mathematische Zeitschrift, vol. 28 (1928), pp. 451–470; *Upon a theory of infinite systems of non-linear implicit and differential equations*, American Journal of Mathematics, vol. 53 (1931), pp. 241–257.

L. Lichtenstein, *Zur Theorie partieller Differentialgleichungen zweiter Ordnung vom hyperbolischen Typus*, Journal für die reine und angewandte Mathematik, vol. 158 (1927), pp. 80–91.

W. T. Reid, *Properties of solutions of an infinite system of ordinary linear differential equations of the first order with auxiliary boundary conditions*, these Transactions, vol. 32 (1930), pp. 284–318.

M. R. Siddiqi, *Zur Theorie der nichtlinearen partiellen Differentialgleichungen vom parabolischen Typus*, Mathematische Zeitschrift, vol. 35 (1932), pp. 464–484.

In addition to the above papers there is also an extensive literature dealing with the single differential equation of infinite order in one unknown function. This theory is closely connected with the Heaviside operational calculus and has little or nothing in common with the theory of infinite systems of differential equations. The device whereby the differential equation

$$F\left(t, x, \frac{dx}{dt}, \frac{d^2x}{dt^2}, \cdots, \frac{d^nx}{dt^n}\right) = 0$$

can be put into the form of a system of equations

$$\frac{dx_k}{dt} = f_k(t, x_1, x_2, \cdots, x_n) \qquad (k = 1, 2, \cdots, n),$$

by setting

$$x_k = \frac{d^{k-1}x}{dt^{k-1}},$$

apparently fails when $n=\infty$. The bibliography of the single equation of infinite order may be found in a footnote to a paper by H. T. Davis in the Annals of Mathematics, (2), vol. 32 (1931), pp. 686–714. It mentions the following authors: Bourlet, Bromwich, von Koch, Pincherle, Ritt, Schürer, Scheffer, Valiron, Wiener.

Part I

1. Notation, terminology, definitions, and lemmas. We consider infinite systems of equations of the form,

$$x_k(t) = \phi_k(t) + \frac{1}{\mu_k}\int_0^t f_k\{\tau, x(\tau)\} \sin \mu_k(t-\tau)d\tau \quad (k = 1, 2, \cdots). \tag{1.1}$$

The $x_k(t)$ are the unknown functions.

The μ_k are any positive numbers.

$\phi_k(t)$ is an abbreviation for $a_k \cos \mu_k t + (a_k'/\mu_k) \sin \mu_k t$, where a_k and a_k' are for the present completely arbitrary, except that, in common with all other numbers arising in this paper, they are real.

$f_k\{\tau, x(\tau)\}$ is a function depending upon k, τ, $x_1(\tau)$, $(d/d\tau)\, x_1(\tau)$, $x_2(\tau)$, $(d/d\tau)\, x_2(\tau)$, $\cdots$.

In general, an italic letter followed by $\{t, x\}$ will be an abbreviation for a function dependent upon the infinitely many independent variables, t, x_1, x_1', x_2, x_2', x_3, x_3', $\cdots$. On the other hand, a Greek letter, with a superscript n, followed by $\{t, x\}$ will indicate a function of the first $2n+1$ of these variables

Thus $F\{t, x\}$ depends upon t, x_1, x_1', x_2, x_2', $\cdots$, while $\psi^{(n)}\{t, x\}$ depends upon t, x_1, x_1', $\cdots$, x_n, x_n' only.

By a "point" in "function space" we shall mean an infinite sequence of numbers, called "coordinates." We shall deal with two types of function space:

In considering type 1, the nth coordinate of a point will usually be denoted by a letter with the subscript n, e.g. x_n. A point in function space of type 1, whose coordinates are represented by x_1, x_2, x_3, $\cdots$, will be denoted briefly by $[x]$.

In dealing with type 2, the nth coordinate will be denoted by a letter unprimed with the subscript $\frac{1}{2}(n+1)$, if n is odd, and primed with the subscript $\frac{1}{2}n$, if n is even. Thus the symbols x_1, x_1', x_2, x_2', x_3, x_3', $\cdots$ may be taken to represent in the proper order the coordinates of a point in space of type 2. Such a point with coordinates represented by these symbols is denoted by (x). Here we use parentheses instead of the square brackets reserved for points of function space of type 1.

We shall use different "distance" functions for the two spaces. We begin by defining the following symbols:

$$|x|_{m,n} = \left[\sum_{k=m}^{n} |x_k|^p\right]^{1/p}; \quad |x| = |x|_{1,\infty} \text{ if this limit exists;}$$

$$\|x\|_{m,n} = \left[\sum_{k=m}^{n} \mu_k^p |x_k|^p + \sum_{k=m}^{n} |x_k'|^p\right]^{1/p}; \quad \|x\| = \|x\|_{1,\infty} \text{ if this limit exists.}$$

p is a positive constant not less than 1. For the sake of brevity, the dependence of these symbols on μ_k and p is not indicated. The "distance" between two points $[b]$ and $[c]$ is defined as $|b-c|$. The "distance" between two points (b) and (c) is defined as $\|b-c\|$.

The symbols obey the following classic inequalities:

$$(1.2)^* \qquad \begin{array}{ll} |b+c|_{m,n} \leqq |b|_{m,n} + |c|_{m,n}; & |b+c| \leqq |b| + |c|; \\ \|b+c\|_{m,n} \leqq \|b\|_{m,n} + \|c\|_{m,n}; & \|b+c\| \leqq \|b\| + \|c\|. \end{array}$$

For our purposes, a region in function space is simply the collection of points whose coordinates satisfy certain conditions. These conditions are usually given in the form of inequalities. Two very special regions Q and R will be largely used in this paper. They are defined as follows:

Let q be a positive number. *The point* (x) *belongs to the region* $Q(q)$ *if*

$$(1.3) \qquad \|x\| \leqq q.$$

Let r be a positive number. *The point* (x) *belongs to the region* $R(r)$ *if for at least one value of* t *the inequality*

$$(1.4) \qquad \|x - \phi(t)\| \leqq r \text{ is valid,}$$

where $\phi_1(t)$, $\phi_2(t)$, $\cdots$ have already been defined and $\phi_k'(t) = (d/dt)\ \phi_k(t)$.

The functions $f_k\{t, x\}$ which we consider are of a special type which we shall call "convergent." A function of this type, depending upon an *infinite* number of variables, is defined as the limit of a sequence of functions, *each one* of which depends only upon a *finite* number of variables. To be more precise, we write a formal definition:

Definition. *A function* $f\{t, x\}$, *defined for* t *in some interval,* $0 \leqq t \leqq T$, *and for* (x) *in some region* S *of (type 2) function space, is said to be of convergent type, if there exists a sequence of functions,* $\psi^{(n)}\{t, x\}$, $n = 1, 2, \cdots$, *the nth function* $\psi^{(n)}\{t, x\}$ *being defined for* $0 \leqq t \leqq T$ *and for all sets of values for* x_1,

* Cf. F. Riesz, *Les Systèmes d'Equations Linéaires à une Infinité d'Inconnues*, p. 43 et seq.

$x_1', x_2, x_2', \cdots, x_n, x_n'$ which are the first $2n$ coordinates of any point (x) in S, such that for any fixed point (x) in S,

$$\lim_{n=\infty} \psi^{(n)}\{t, x\} = f\{t, x\}.$$

The usefulness of this definition rests on the following lemma:

LEMMA 1. *Let $\psi^{(n)}$ be continuous in its $2n+1$ arguments. Let $|\psi^{(n)}\{t, x\}| \leqq M$, where M is some number independent of n, t, x_1, x_1', x_2, x_2', $\cdots$. Let $x_1(t)$, $x_2(t)$, $\cdots$ be a set of functions, each of which is defined and of class C' on $0 \leqq t \leqq T$, and set $x_k'(t) = (d/dt)x_k(t)$. Let these functions be such that $(x(t))$ lies in S for $0 \leqq t \leqq T$. Finally let $g(t)$ be defined and integrable on $0 \leqq t \leqq T$.*

Then $\int_0^T f\{t, x(t)\}\, g(t)dt$ exists in the sense of Lebesgue and is in fact equal to $\lim_{n=\infty} \int_0^T \psi^{(n)}\{t, x(t)\}\, g(t)dt$.

The proof of this lemma together with the following corollary is left to the reader.

COROLLARY. *Let μ be a constant. Then under the hypotheses of Lemma 1 $\int_0^t f\{\tau, x(\tau)\} \sin \mu(t-\tau)d\tau$ is of class C' for $0 \leqq t \leqq T$, possessing almost everywhere in this interval a second derivative.*

We need two more simple inequalities before proceeding to the existence proof of the next section.

If $f(t)$ is integrable on $0 \leqq t \leqq T$, we have the classic inequality

$$\left[\int_0^t |f(\tau)|\, d\tau\right]^p \leqq t^{p-1}\int_0^t |f(\tau)|^p d\tau^* \text{ for } 0 \leqq t \leqq T.$$

Hence, if $f_1(t)$, $f_2(t)$, $\cdots$ is an infinite sequence of functions, each of which is integrable on $0 \leqq t \leqq T$, then

$$\left\| \int_0^t |f(\tau)|\, d\tau \right\|_{m,n}^p \leqq t^{p-1}\int_0^t \| f(\tau) \|_{m,n}^p d\tau \text{ for } 0 \leqq t \leqq T. \tag{1.5}$$

If $\| f(t) \|_{m,n}$ is bounded uniformly with respect to n and t, and if $\| f(t) \|_{m,\infty}$ exists for each t, we can, by Lebesgue's theorem, pass to the limit and write

$$\left\| \int_0^t |f(\tau)|\, d\tau \right\|_{m,\infty}^p \leqq t^{p-1}\int_0^t \| f(\tau) \|_{m,\infty}^p d\tau. \tag{1.6}$$

2. **Fundamental existence theorem for equations (1.1).** We suppose that there exist four positive numbers r, A, B, T, such that the following three hypotheses hold:

---

* E. W. Hobson, *Functions of a Real Variable*, 3d edition, vol. 1, p. 643.

Hypothesis 2.I. *The $f_k\{t, x\}$ are defined and are of convergent type for (x) in $R(r)$ and for $0 \leqq t \leqq T$. The approximation functions $\psi_k^{(n)}\{t, x\}$ are continuous and uniformly bounded for each k.*

Hypothesis 2.II. $|f\{t, \phi(t')\}| \leqq B$ *for* $0 \leqq t \leqq T$ *and* $-\infty < t' < +\infty$.

Hypothesis 2.III. $|f\{t, x\} - f\{t, \bar{x}\}| \leqq A \cdot \|x - \bar{x}\|$, *where* (x) *and* $(\bar{x})$ *are both points of* $R(r)$ *and* $0 \leqq t \leqq T$.

Then there exists a unique set of functions $x_1(t), x_2(t), \cdots$, each of class C' on the interval $0 \leqq t \leqq K$ (K is the smaller of the two numbers T and $2^{-1/p}r/(Ar+B)$) with the following two properties:

I. *$(x(t))$ belongs to R for $0 \leqq t \leqq K$.*

II. *If these functions are substituted in* (1.1) *the right hand members exist in the sense of Lebesgue and are identically equal to the left members for* $0 \leqq t \leqq K$.

Such a set of functions will be called a solution.

We first note as a consequence of Hypotheses 2.II and 2.III and (1.2) that

$$|f\{t, x\}| \leqq Ar + B = C \text{ for } (x) \text{ in } R. \tag{2.1}$$

The actual solution is constructed from the following system of successive approximations:

$$\begin{aligned} x_k^{(0)}(t) &= \phi_k(t), \cdots, \\ x_k^{(n)}(t) &= \phi_k(t) + \frac{1}{\mu_k}\int_0^t f_k\{\tau, x^{(n-1)}(\tau)\} \sin \mu_k(t-\tau)d\tau \\ &\qquad\qquad (n = 1, 2, 3, \cdots). \end{aligned} \tag{2.2}$$

Differentiating these, we have also

$$\begin{aligned} x_k^{(0)\prime}(t) &= \phi_k'(t), \cdots, \\ x_k^{(n)\prime}(t) &= \phi_k'(t) + \int_0^t f_k\{\tau, x^{(n-1)}(\tau)\} \cos \mu_k(t-\tau)d\tau. \end{aligned} \tag{2.3}$$

We prove by induction that, for $0 \leqq t \leqq K$, $x_k^{(n)}(t)$ exists and is of class C' (cf. Lemma I and its corollary), and that $(x^{(n)}(t))$ belongs to R. Assuming these facts true for $(x^{(n-1)}(t))$, it follows from (2.2), (2.3), and (1.6) that

$$\begin{aligned} \|x^{(n)}(t) - \phi(t)\|^p = &\left|\int_0^t f\{\tau, x^{(n-1)}(\tau)\} \sin \mu(t-\tau)d\tau\right|^p \\ &+ \left|\int_0^t f\{\tau, x^{(n-1)}(\tau)\} \cos \mu(t-\tau)d\tau\right|^p \end{aligned}$$

$$\leqq 2\left|\int_0^t |f\{\tau, x^{(n-1)}(\tau)\}|\,d\tau\right|^p \leqq 2t^{p-1}\int_0^t \|f\{\tau, x^{(n-1)}(\tau)\}\|^p d\tau$$

$$\leqq 2t^{p-1}\int_0^t C^p d\tau = 2C^p t^p \leqq r^p, \text{ i.e. } (x^{(n)}(t)) \text{ belongs to } R.$$

Since the stated facts are obviously true for $x_k^{(0)}(t)$, they are by induction true for $x_k^{(n)}(t)$.

We next prove that $x_k^{(n)}(t)$ converges uniformly toward a limit function $x_k(t)$ as n becomes infinite. By setting $n=1$, we have from the above inequalities

$$\|x^{(1)}(t) - x^{(0)}(t)\|^p = \|x^{(1)}(t) - \phi(t)\|^p \leqq 2C^p t^p.$$

We also obtain from (2.2), (2.3), (1.6), and Hypothesis 2.III

$$\|x^{(n+1)}(t) - x^{(n)}(t)\|^p = \left|\int_0^t [f\{\tau, x^{(n)}(\tau)\} - f\{\tau, x^{(n-1)}(\tau)\}]\sin\mu(t-\tau)d\tau\right|^p$$

$$+\left|\int_0^t [f\{\tau, x^{(n)}(\tau)\} - f\{\tau, x^{(n-1)}(\tau)\}]\cos\mu(t-\tau)d\tau\right|^p$$

$$\leqq 2\left|\int_0^t |f\{\tau, x^{(n)}(\tau)\} - f\{\tau, x^{(n-1)}(\tau)\}|\,d\tau\right|^p$$

$$\leqq 2t^{p-1}\int_0^t \|f\{\tau, x^{(n)}(\tau)\} - f\{\tau, x^{(n-1)}(\tau)\}\|^p d\tau$$

$$\leqq 2A^p t^{p-1}\int_0^t \|x^{(n)}(\tau) - x^{(n-1)}(\tau)\|^p d\tau.$$

It then follows by induction that

$$\|x^{(n)}(t) - x^{(n-1)}(t)\|^p \leqq \frac{2^n C^p A^{p(n-1)} t^{np}}{(p+1)(2p+1)\cdots([n-1]p+1)}.$$

The uniform convergence of $x_k^{(n)}(t)$ and $x_k^{(n)\prime}(t)$ for $0 \leqq t \leqq K$ now follows from the Weierstrass test.

We also find, using (1.2), that

$$\|x(t) - x^{(n)}(t)\| \leqq \sum_{m=n}^{\infty}\|x^{(m+1)}(t) - x^{(m)}(t)\|$$

$$\leqq \sum_{m=n}^{\infty}\frac{2^{(m+1)/p}CA^m K^{m+1}}{[(p+1)(2p+1)\cdots(mp+1)]^{1/p}},$$

which is the remainder after $(n-1)$ terms of a certain convergent series of

positive constants. Now $|f\{t, x(t)\} - f\{t, x^{(n)}(t)\}| \leqq A \cdot \|x(t) - x^{(n)}(t)\|$, so that

$$\lim_{n=\infty} f_k\{t, x^{(n)}(t)\} = f_k\{t, x(t)\}$$

uniformly for $0 \leqq t \leqq K$.

This is all that is needed to complete the proof that the functions $x_k(t)$ satisfy equations (1.1). The proof of the uniqueness of this solution follows essentially the same lines and is left to the reader.

3. **Approximation by solutions of finite systems.** Let n be a positive integer. Then corresponding to the infinite system of equations (1.1) there is also the finite system,

$$(3.1)\quad x_{nk}(t) = \phi_k(t) + \frac{1}{\mu_k}\int_0^t \psi_k^{(n)}\{\tau, x_n(\tau)\} \sin \mu_k(t-\tau)d\tau \quad (k = 1, 2, \cdots, n)$$

for determining the unknown functions $x_{nk}(t)$. It is the main object of this article to consider, under certain hypotheses, the approximation to $x_1(t)$, $\cdots$, $x_n(t)$, which are the first n functions of the solution of (1.1), by the functions $x_{n1}(t), x_{n2}(t), \cdots, x_{nn}(t)$, which form the solution of (3.1). We prove the following

THEOREM. *Suppose that the limits $|a\mu|$ and $|a'|$ exist, and that there are four positive numbers r, A, B, and T independent of n so that the following hypotheses hold*:

HYPOTHESIS 3.I. *The convergent functions $f_k\{t, x\}$ are defined in a region $Q(q)$, where $q = r + 2^{1/p}(|a\mu| + |a'|)$, and for $0 \leqq t \leqq T$* (see (1.3)). *The approximation functions $\psi_k^{(n)}$ have the special form $\psi_k^{(n)}\{t, x\} \equiv f_k(t, x_1, x_1', \cdots, x_n, x_n', 0, 0, 0, \cdots)$, $k = 1, 2, \cdots$.*

HYPOTHESIS 3.II. $|\psi^{(n)}\{t, \phi(t')\}| \leqq B$ *for* $0 \leqq t \leqq T$, $-\infty < t' < +\infty$.

HYPOTHESIS 3.III. $|f\{t,x\} - f\{t, \bar{x}\}| \leqq A \cdot \|x - \bar{x}\|$ *for* $0 \leqq t \leqq T$ *and for* (x) *and* $(\bar{x})$ *in* $Q(q)$; $f_k\{t, x\}$ *is continuous in* t.

Then we may draw the following four conclusions:

CONCLUSION I. $R(r) < Q(q)$.

CONCLUSION II. *Hypotheses* 2.I, 2.II, 2.III *of the preceding article hold and consequently a unique solution, $x_1(t), x_2(t), \cdots$, of* (1.1) *exists for* $0 \leqq t \leqq K$.

CONCLUSION III. *A solution $x_{n1}(t), \cdots, x_{nn}(t)$ of equations* (3.1) *also exists for $0 \leqq t \leqq K$, such that $\|x_n(t) - \phi(t)\|_{1,n} \leqq r$ and $|\psi^{(n)}\{t, x_n(t)\}| \leqq Ar + B = C$.*

CONCLUSION IV. $\lim_{n=\infty} \|x(t) - x_n(t)\|_{1,n} = 0$, $\lim_{n=\infty} \|x(t)\|_{n+1,\infty} = 0$, $\lim_{n=\infty} \int_0^T |\psi^{(n)}\{t, x_n(t)\}|_{n+1,\infty}^p\, dt = 0$, *the first two of these limits holding uniformly for* $0 \leqq t \leqq K$.

Conclusions I, II, and III are sufficiently obvious to require no proof. We content ourselves with a proof of IV.

From Hypotheses 3.I and 3.III we have

$$\begin{aligned}(3.2)\qquad \|f\{t, x\} - \psi^{(n)}\{t, \bar{x}\}\|^p \leqq A^p\|x - \bar{x}\|_{1,n}^p \\ + A^p\|x\|_{n+1,\infty}^p \text{ if } (x) \text{ and } (\bar{x}) \text{ both belong to } Q.\end{aligned}$$

Let ϵ be a preassigned positive number arbitrarily small. Choose N_1 so large that

$$(3.3)\qquad \|a\mu\|_{n+1,\infty} + \|a'\|_{n+1,\infty} \leqq \epsilon \text{ for } n \geqq N_1.$$

Inequality (2.1) holds, and it is easy to justify the relation

$$\int_0^K \|f\{\tau, x(\tau)\}\|^p d\tau = \sum_{k=1}^{\infty} \int_0^K | f_k\{\tau, x(\tau)\} |^p d\tau \leqq C^p K \qquad (C = Ar + B).$$

Consequently it is possible to find N_2 so great that

$$(3.4)\qquad \int_0^K \|f\{\tau, x(\tau)\}\|_{n+1,\infty}^p d\tau = \sum_{k=n+1}^{\infty} \int_0^K | f_k\{\tau, x(\tau)\} |^p d\tau < \epsilon \text{ for } n \geqq N_2.$$

Let N_3 be the greater of the two numbers N_1 and N_2 so that both (3.3) and (3.4) hold as long as $n \geqq N_3$. Now, from (1.1) and (1.6), we have

$$\begin{aligned}\|x(t) - \phi(t)\|_{n+1,\infty}^p &\leqq 2\left\| \int_0^t | f\{\tau, x(\tau)\} | \, d\tau \right\|_{n+1,\infty}^p \\ &\leqq 2K^{p-1} \int_0^K \|f\{\tau, x(\tau)\}\|_{n+1,\infty}^p d\tau \leqq 2K^{p-1}\epsilon.\end{aligned}$$

Also

$$\|x(t)\|_{n+1,\infty} \leqq \|x(t) - \phi(t)\|_{n+1,\infty} + \|\phi(t)\|_{n+1,\infty} \leqq (2K^{p-1}\epsilon)^{1/p} + \|\phi(t)\|_{n+1,\infty}.$$

But since it is easy to show that $\|\phi(t)\|_{n+1,\infty} \leqq 2^{1/p} (\|a\mu\|_{n+1,\infty} + \|a'\|_{n+1,\infty}) \leqq 2^{1/p}\epsilon$ we have $\|x(t)\|_{n+1,\infty} \leqq (2K^{p-1}\epsilon)^{1/p} + 2^{1/p}\epsilon$ as long as $n \geqq N_3$.

Since N_3 is independent of t for $0 \leqq t \leqq K$, we have proved the second relation under IV.

For convenience choose a number N_4 so large that

$$(3.5)\qquad \|x(t)\|_{n+1,\infty}^p \leqq \epsilon \text{ as long as } n \geqq N_4.$$

Now set up the successive approximations for equations (3.1):

$$\begin{aligned}\overset{(0)}{x}_{nk}(t) &= \phi_k(t), \cdots, \\ \overset{(m)}{x}_{nk}(t) &= \phi_k(t) + \frac{1}{\mu_k} \int_0^t \psi_k^{(n)}\{\tau, \overset{(m-1)}{x_n}(\tau)\} \sin \mu_k(t - \tau) d\tau, \cdots.\end{aligned}$$

We find now from (1.1) and (1.5) that

$$\|x(t) - x_n^{(m)}(t)\|_{1,n}^p \leqq 2t^{p-1} \int_0^t \|f\{\tau, x(\tau)\} - \psi^{(n)}\{\tau, x_n^{(m-1)}(\tau)\}\|_{1,n}^p d\tau,$$

and this by (3.2) is less than or equal to

$$2t^{p-1} \int_0^t A^p \cdot \|x(\tau) - x_n^{(m-1)}(\tau)\|_{1,n}^p d\tau + 2t^{p-1} \int_0^t A^p \cdot \|x(\tau)\|_{n+1,\infty}^p d\tau.$$

Therefore we obtain the following inequality:

$$(3.6) \quad \|x(t) - x_n^{(m)}(t)\|_{1,n}^p \leqq 2A^p t^{p-1} \int_0^t \|x(\tau) - x_n^{(m-1)}(\tau)\|_{1,n}^p d\tau + 2A^p t^p \epsilon,$$

which is valid for $0 \leqq t \leqq K$ and $n \geqq N_4$. Furthermore it is clear that

$$\|x(\tau) - x_n^{(0)}(\tau)\|_{1,n}^p = \|x(\tau) - \phi(\tau)\|_{1,n}^p \leqq r^p.$$

Hence setting $m = 1$ in (3.6) we obtain

$$\|x(t) - x_n^{(1)}(t)\|_{1,n}^p \leqq 2A^p r^p t^p + \epsilon \cdot 2A^p t^p.$$

It is now easy to prove by induction that

$$\|x(t) - x_n^{(m)}(t)\|_{1,n}^p \leqq \frac{2^m A^{mp} r^p t^{mp}}{(p+1)(2p+1)\cdots([m-1]p+1)}$$
$$+ \epsilon \left[\sum_{j=1}^{m} \frac{2^j A^{jp} t^{jp}}{(0 \cdot p + 1)(1 \cdot p + 1)(2 \cdot p + 1) \cdots ([j-1] \cdot p + 1)} \right].$$

Since $0 \leqq t \leqq K$, we surely must have now

$$\|x(t) - x_n^{(m)}(t)\|_{1,n}^p \leqq \frac{2^m A^{mp} r^p K^{mp}}{(p+1)(2p+1)\cdots([m-1]p+1)} + \epsilon D,$$

where D is equal to the value of the convergent series of positive constants

$$\sum_{j=1}^{\infty} \frac{2^j A^{jp} K^{jp}}{(0 \cdot p + 1)(1 \cdot p + 1)(2 \cdot p + 1) \cdots ([j-1] \cdot p + 1)}.$$

Also it is known that

$$\lim_{m=\infty} x_{nk}^{(m)}(t) = x_{nk}(t)$$

uniformly, and the limit of the first term on the right, as m increases indefinitely, is zero.

Hence $\|x(t)-x_n(t)\|_{1,n}\leqq(\epsilon D)^{1/p}$, as long as $n\geqq N_4$. This proves the first relation under IV.

By (1.2) we have

$$\begin{aligned}\int_0^K \|\psi^{(n)}\{\tau, x_n(\tau)\}\|_{n+1,\infty}d\tau &\leqq \int_0^K \|f\{\tau, x(\tau)\}\|_{n+1,\infty}d\tau \\ &+ \int_0^K \|\psi^{(n)}\{\tau, x_n(\tau)\} - f\{\tau, x(\tau)\}\|_{n+1,\infty}d\tau.\end{aligned} \tag{3.7}$$

We first appraise the first term on the right:

$$\begin{aligned}\left[\int_0^K \|f\{\tau, x(\tau)\}\|_{n+1,\infty}d\tau\right]^p &\leqq K^{p-1}\int_0^K \|f\{\tau, x(\tau)\}\|^p_{n+1,\infty}d\tau \\ &\leqq \epsilon K^{p-1} \text{ by (3.4) for } n \geqq N_3.\end{aligned}$$

Thus the first term on the right of (3.7) is not greater than $(\epsilon K^{p-1})^{1/p}$ for $n\geqq N_3$. The second term on the right of (3.7) is appraised by means of (3.2):

$$\begin{aligned}\|\psi^{(n)}\{\tau, x_n(\tau)\} - f\{\tau, x(\tau)\}\|^p_{n,\infty} &\leqq A^p\cdot\|x(\tau) - x_n(\tau)\|^p_{1,n} + A^p\cdot\|x(\tau)\|^p_{n+1,\infty} \\ &\leqq \epsilon A^p(D+1) \text{ for } n \geqq N_4.\end{aligned}$$

Thus the second term on the right of (3.7) is not greater than $KA(\epsilon(D+1))^{1/p}$ for $n\geqq N_4$. Therefore

$$\int_0^K \|\psi^{(n)}\{\tau, x_n(\tau)\}\|_{n+1,\infty}d\tau \leqq \frac{E\epsilon^{1/p}}{C^{p-1}}, \text{ for } n \geqq N_5,$$

where N_5 is the greater of the two numbers N_3 and N_4, where $C=Ar+B$, and where $E=C^{p-1}[KA(D+1)^{1/p}+K^{(p-1)/p}]$. Also, from the obvious relations

$$\|\psi^{(n)}\|^p_{n+1,\infty} = \|\psi^{(n)}\|^{p-1}_{n+1,\infty}\cdot\|\psi^{(n)}\|_{n+1,\infty} \leqq C^{p-1}\cdot\|\psi^{(n)}\|_{n+1,\infty},$$

it follows that

$$\int_0^K \|\psi^{(n)}\{\tau, x_n(\tau)\}\|^p_{n+1,\infty}d\tau \leqq E\epsilon^{1/p}, \text{ as long as } n \geqq N_5.$$

This completes the proof of the theorem.

Part II

4. The application of the results of Part I to partial differential equations. We consider partial differential equations of the form

$$P(u) \equiv \frac{\partial^2 u}{\partial t^2} - \frac{\partial^2 u}{\partial y^2} - F\left(\frac{\partial u}{\partial y}, \frac{\partial u}{\partial t}, u, y, t\right) = 0, \tag{4.1}$$

where $F(p_1, p_2, u, y, t)$ is defined for $-\infty < p_1 < +\infty, -\infty < p_2 < +\infty, |u| \leqq h$, $0 \leqq y \leqq \pi$, $0 \leqq t \leqq T$; is uniformly continuous in y and t; and obeys a Lipschitz condition in p_1, p_2, and u:

$$\begin{aligned} &|F(p_1, p_2, u, y, t) - F(\bar{p}_1, \bar{p}_2, \bar{u}, y, t)| \\ &\qquad \leqq \alpha \cdot |p_1 - \bar{p}_1| + \beta \cdot |p_2 - \bar{p}_2| + \gamma \cdot |u - \bar{u}|, \end{aligned} \tag{4.2}$$

as long as $|u| \leqq h$, $|\bar{u}| \leqq h$.

We retain the notation of §1 with the understanding, however, that $p = 2$ and $\mu_k = k$.

Let

$$q = \frac{h}{\left\|\frac{1}{\mu}\right\|} = \frac{h}{\left(\sum_{k=1}^{\infty} \frac{1}{k^2}\right)^{1/2}} = \frac{6^{1/2} h}{\pi}.$$

Let there be given two functions, $f(y)$ and $g(y)$, defined for the interval $0 \leqq y \leqq \pi$ and subject to the following conditions:

$f(0) = f(\pi) = 0$. $f(y)$ is an indefinite integral possessing a derivative whose square is summable, and furthermore

$$\frac{2}{\pi} \int_0^\pi [f'(y)]^2 dy < \tfrac{1}{8} q^2.$$

$g(y)$ has a summable square, and furthermore

$$\frac{2}{\pi} \int_0^\pi [g(y)]^2 dy < \tfrac{1}{8} q^2.$$

We shall try to find a solution of (4.1) such that

$$u(0, t) = u(\pi, t) = 0, \quad u(y, 0) = f(y), \quad \left.\frac{\partial u}{\partial t}\right|_{t=0} = g(y). \tag{4.3}$$

We know from the theory of trigonometric series that

$$\frac{2}{\pi} \int_0^\pi [f'(y)]^2 dy = \|a\mu\|^2 < \tfrac{1}{8} q^2 \text{ where } a_k \mu_k = a_k k = \frac{2}{\pi} \int_0^\pi f'(y) \cos ky \, dy$$

and

$$\frac{2}{\pi} \int_0^\pi [g(y)]^2 dy = \|a'\|^2 < \tfrac{1}{8} q^2 \text{ where } a_k' = \frac{2}{\pi} \int_0^\pi g(y) \sin ky \, dy.$$

Remembering that $f(0) = f(\pi) = 0$, we find on integrating by parts that

$$a_k = \frac{2}{\pi} \int_0^\pi f(y) \sin ky \, dy.$$

We also have the obvious inequality $2^{1/2}(\|a\mu\|+\|a'\|)<q$. Let $r=q-2^{1/2}(\|a\mu\|+\|a'\|)$.

Let (x) be a point of $Q(q)$. Then from Schwarz's inequality we have

$$\sum_{k=1}^{\infty}|x_k| \leqq \|x\mu\|\cdot\left\|\frac{1}{\mu}\right\| < \|x\|\cdot\left\|\frac{1}{\mu}\right\| \leqq q\cdot\frac{\pi}{6^{1/2}} = h.$$

Hence the series $u(y)=\sum_{k=1}^{\infty} x_k \sin ky$ converges absolutely and uniformly with respect to y. With the help of the Riesz-Fischer theorem we may now make the following statement:

Corresponding to a point (x) in Q there is defined a continuous function, $u(y)$, for the interval $0\leqq y\leqq\pi$, possessing almost everywhere a derivative $u'(y)$, whose square is summable, such that

$$\mu_k x_k = \frac{2}{\pi}\int_0^{\pi} u'(y)\cos ky\,dy,\ u(y) = \int_0^{y} u'(\eta)d\eta,\, u(0) = u(\pi) = 0,\ |u(y)| < h.$$

Corresponding to this same point (x) there is also defined for this interval a second function $v(y)$ whose square is summable and such that

$$x_k' = \frac{2}{\pi}\int_0^{\pi} v(y)\sin ky\,dy.$$

$u'(y)$ and $v(y)$ are unique on $0\leqq y\leqq\pi$ except possibly for point sets of measure zero.

Now let

$$f_k\{t, x\} = \frac{2}{\pi}\int_0^{\pi} F(u'(y), v(y), u(y), y, t)\sin ky\,dy. \tag{4.4}$$

Let $[u(y), v(y)]$ and $[\bar{u}(y), \bar{v}(y)]$ be the two pairs of functions corresponding respectively to the two points (x) and $(\bar{x})$ in Q. From (4.2) we have

$$\begin{aligned}&[F(u'(y), v(y), u(y), y, t) - F(\bar{u}'(y), \bar{v}(y), \bar{u}(y), y, t)]^2\\ &\qquad \leqq 3\alpha^2\cdot[u'(y)-\bar{u}'(y)]^2 + 3\beta^2\cdot[v(y)-\bar{v}(y)]^2 + 3\gamma^2\cdot[u(y)-\bar{u}(y)]^2.\end{aligned}$$

Now the $f_k\{t, x\}$ are the Fourier coefficients of $F(u'(y), v(y), u(y), y, t)$ while the $f_k\{t, \bar{x}\}$ are the Fourier coefficients of $F(\bar{u}'(y), \bar{v}(y), \bar{u}(y), y, t)$, and hence the $[f_k\{t, x\} - f_k\{t, \bar{x}\}]$ are the Fourier coefficients of

$$[F(u'(y), v(y), u(y), y, t) - F(\bar{u}'(y), \bar{v}(y), \bar{u}(y), y, t)].$$

Hence

$$\|f\{t, x\} - f\{t, \bar{x}\}\|^2 = \frac{2}{\pi}\int_0^\pi [F(u'(y), v(y), u(y), y, t)$$
$$- F(\bar{u}'(y), \bar{v}(y), \bar{u}(y), y, t)]^2 dy$$
$$\leqq \frac{6\alpha^2}{\pi}\int_0^\pi [u'(y) - \bar{u}'(y)]^2 dy + \frac{6\beta^2}{\pi}\int_0^\pi [v(y) - \bar{v}(y)]^2 dy$$
$$+ \frac{6\gamma^2}{\pi}\int_0^\pi [u(y) - \bar{u}(y)]^2 dy$$
$$= 3\alpha^2 \cdot \|(x - \bar{x})\mu\|^2 + 3\beta^2 \cdot \|x' - \bar{x}'\|^2 + 3\gamma^2 \cdot \|x - \bar{x}\|^2$$
$$\leqq (3\alpha^2 + 3\gamma^2) \cdot \|(x - \bar{x})\mu\|^2 + 3\beta^2 \cdot \|x' - \bar{x}'\|^2$$
$$\leqq A^2 \cdot \|x - \bar{x}\|^2,$$

where A^2 is the greater of the two numbers $(3\alpha^2+3\gamma^2)$ and $3\beta^2$. Thus we have

$$\|f\{t, x\} - f\{t, \bar{x}\}\| \leqq A \cdot \|x - \bar{x}\|. \tag{4.5}$$

Also $\|f\{t, x\}\| \leqq \|f\{t, 0\}\| + A \cdot \|x\|$, which is obviously bounded for (x) in Q. Finally we write

$$\psi_k^{(n)}\{t, x\} \equiv f_k(t, x_1, x_1', \cdots, x_n, x_n', 0, 0, 0, 0, \cdots), \tag{4.6}$$

from which it follows, using (4.5), that

$$\lim_{n=\infty} \psi_k^{(n)}\{t, x\} = f_k\{t, x\}.$$

Thus we easily see that the $f_k\{t, x\}$ as here defined satisfy Hypotheses 3.I, 3.II, 3.III. Hence all the results of Part I are now available.

It will be necessary to generalize our idea of a "solution" of a partial differential equation. We first make the following

Definition. *A continuous function, $u(y, t)$, defined on the rectangular region $0 \leqq y \leqq \pi$, $0 \leqq t \leqq K$, and possessing first partial derivatives almost everywhere in this rectangle, is a solution in the generalized sense of the second-order partial differential equation $P(u) = 0$, if there exists a sequence of functions, $u_1(y, t)$, $u_2(y, t)$, $u_3(y, t)$, $\cdots$, each of class C'' in this same rectangle, such that the following four conditions hold:*

(I) $\lim_{n=\infty} u_n(y, t) = u(y, t)$ *uniformly in y and t;*

(II) $\lim_{n=\infty} \int_0^\pi \left[\frac{\partial u}{\partial y} - \frac{\partial u_n}{\partial y}\right]^2 dy = 0$ *uniformly in t;*

(III) $\lim_{n=\infty} \int_0^\pi \left[\frac{\partial u}{\partial t} - \frac{\partial u_n}{\partial t}\right]^2 dy = 0$ *uniformly in t;*

(IV) $$\lim_{n\to\infty} \int_0^K \int_0^\pi [P(u_n(y, t))]^2 dy\, dt = 0.$$

In the next article there will be given a more general definition, which, however, for our present purposes is equivalent to the definition above.

Let $(x(t))$ be a solution, valid for $0 \leqq t \leqq K$, of the infinite system of equations (1.1). We know from §2 that such a solution exists and is in fact unique. *Then the function*

$$u(y, t) = \sum_{k=1}^{\infty} x_k(t) \sin ky \tag{4.7}$$

is a solution (in the generalized sense) of (4.1) *and satisfies the conditions* (4.3). The approximation functions, $u_n(y, t)$, mentioned in the definition will be provided for as follows:

$$u_n(y, t) = \sum_{k=1}^{n} x_{nk}(t) \sin ky, \tag{4.8}$$

where $x_{n1}(t), \cdots, x_{nn}(t)$ satisfy the finite system (3.1). Since there is no difficulty about differentiating the finite sum in (4.8), we see that $u_n(y, t)$ satisfies the requirement of being of class C''.

It is intuitively evident that $u(y, t)$ possesses almost everywhere the partial derivatives

$$\frac{\partial u}{\partial t} = \sum_{k=1}^{\infty} x_k'(t) \sin ky \text{ and } \frac{\partial u}{\partial y} = \sum_{k=1}^{\infty} x_k(t) k \cos ky,$$

where the indicated series converge in the mean on the interval $0 \leqq y \leqq \pi$ for each t. They do not necessarily converge in the usual sense. The rigorous proof of these facts is omitted because it merely involves some of the fundamental classical analysis concerning double limits and convergence in the mean.

We notice also that $u(y, t)$ and $\partial u/\partial t$ take on the preassigned initial values of (4.3). For $x_k(0) = a_k$ and $x_k'(0) = a_k'$.

It remains to show that $u(y, t)$ and the $u_n(y, t)$ satisfy the Conditions I, II, III, IV of the definition.

Proof that Condition I holds. Let ϵ be an arbitrarily small positive number. We have

$$\left| u(y, t) - u_n(y, t) \right| = \left| \sum_{k=1}^{n} [x_k(t) - x_{nk}(t)] \sin ky + \sum_{k=n+1}^{\infty} x_k(t) \sin ky \right|$$
$$\leqq \sum_{k=1}^{n} \left| x_k(t) - x_{nk}(t) \right| + \sum_{k=n+1}^{\infty} \left| x_k(t) \right|.$$

Using Schwarz's inequaity we find that

$$\sum_{k=1}^{n} | x_k(t) - x_{nk}(t) | \leqq \left|[x(t) - x_{nk}(t)]\mu\right|_{1,n} \cdot \left|\frac{1}{\mu}\right| \leqq \|x(t) - x_n(t)\|_{1,n} \cdot \frac{\pi}{6^{1/2}},$$

and this may be taken less than $\frac{1}{2}\epsilon$, according to the theorem of §3, if n is chosen greater than some number N_1, independent of t. Likewise from Schwarz's inequality and the theorem of §3 we have

$$\sum_{k=n+1}^{\infty} | x_k(t) | \leqq \left|x(t)\mu\right|_{n+1,\infty} \cdot \left|\frac{1}{\mu}\right| \leqq \|x(t)\|_{n+1,\infty} \cdot \frac{\pi}{6^{1/2}} \leqq \frac{1}{2}\epsilon$$

if n is greater than some number N_2, independent of t. Hence $|u(y, t) - u_n(y, t)| < \epsilon$ for $n > N_3$, where N_3 is the greater of the two numbers N_1 and N_2.

Proof that Condition II holds. The function $(\partial u/\partial y - \partial u_n/\partial y)$ has the Fourier coefficients $(x_k(t) - x_{nk}(t))k$, for $k = 1, 2, \cdots, n$, and $x_k(t)k$, for $k = n+1, n+2, n+3, \cdots$. Consequently

$$\frac{2}{\pi}\int_0^{\pi}\left(\frac{\partial u}{\partial y} - \frac{\partial u_n}{\partial y}\right)^2 dy = \left|[x(t) - x_n(t)]\mu\right|_{1,n}^2 + \left|x(t)\mu\right|_{n+1,\infty}^2$$

$$\leqq \|x(t) - x_n(t)\|_{1,n}^2 + \|x(t)\|_{n+1,\infty}^2$$

and this by §3 converges uniformly to zero.

Proof that Condition III holds. The function $(\partial u/\partial t - \partial u_n/\partial t)$ has the Fourier coefficients $x_k'(t) - x_{nk}'(t)$, for $k = 1, \cdots, n$, and $x_k'(t)$, for $k = n+1, n+2, \cdots$. Consequently

$$\frac{2}{\pi}\int_0^{\pi}\left(\frac{\partial u}{\partial t} - \frac{\partial u_n}{\partial t}\right)^2 dy = \left|x'(t) - x_n'(t)\right|_{1,n}^2 + \left|x'(t)\right|_{n+1,\infty}^2$$

$$\leqq \|x(t) - x_n(t)\|_{1,n}^2 + \|x(t)\|_{n+1,\infty}^2,$$

which converges uniformly to zero as above.

Proof that Condition IV holds. Differentiate $u_n(y, t)$ and substitute in the operator, $P(\ )$. $P[u_n(y, t)]$ thus is a function of y and t. It is found from (4.1), (4.4), and (4.6) that

$$\frac{2}{\pi}\int_0^{\pi} P[u_n(y, t)] \sin ky\, dy = x_{nk}''(t) + k^2 x_{nk} - \psi_k^{(n)}\{t, x_n(t)\},$$

$$\text{for } k = 1, 2, \cdots, n,$$

and this vanishes on account of (3.1). Also

$$\frac{2}{\pi}\int_0^{\pi} P[u_n(y, t)] \sin ky\, dy = -\psi_k^{(n)}\{t, x_n(t)\},$$

$$\text{for } k = n + 1, n + 2, n + 3, \cdots.$$

Hence, by Parseval's theorem,

$$\frac{2}{\pi}\int_0^\pi \{P[u_n(y,t)]\}^2 dy = \|\psi^{(n)}\{t, x_n(t)\}\|_{n+1,\infty}^2,$$

and integrating this with respect to t between the limits 0 and K we get a quantity which, by the theorem of §3, may be taken arbitrarily small by taking n sufficiently large.

5. **A study of the conception of a "solution in the generalized sense" with special reference to partial differential equations of the form (4.1).** It will be shown in this section that a solution, $u(y, t)$, of (4.1) in the generalized sense is also a solution in the ordinary sense, provided that u possesses second derivatives and F satisfies certain simple requirements as to continuity and differentiability. It will also be shown that the solution in the generalized sense obtained in the previous section is the only such solution which satisfies the boundary conditions (4.3).

It is easy to prove these theorems for equation (4.1) because of its especially simple structure. But the generalized notion of a solution of a partial differential equation is naturally of a much broader character. It might very well prove useful in the treatment of all partial differential equations, especially those of the hyperbolic type. I give here a complete definition, slightly more general than the one introduced in §4, which was not symmetrical in y and t.

DEFINITION 1. *A continuous function, $u(y, t)$, defined in some finite region E and possessing almost everywhere in E first partial derivatives, is a solution in the generalized sense of the second-order partial differential equation $P(u)=0$ if there exists a sequence of functions, $u_1(y, t)$, $u_2(y, t)$, $u_3(y, t)$, $\cdots$, each defined in E and each of class C'', such that the following four conditions hold*:

$$\text{(I)} \qquad \lim_{n=\infty} u_n(y,t) = u(y,t) \text{ uniformly in } E;$$

$$\text{(II)} \qquad \lim_{n=\infty} \iint_E \left| \frac{\partial u}{\partial y} - \frac{\partial u_n}{\partial y} \right| dy\,dt = 0;$$

$$\text{(III)} \qquad \lim_{n=\infty} \iint_E \left| \frac{\partial u}{\partial t} - \frac{\partial u_n}{\partial t} \right| dy\,dt = 0;$$

$$\text{(IV)} \qquad \lim_{n=\infty} \iint_E | P(u_n) | \, dy\,dt = 0.$$

In the sequel, E will be assumed to be a closed simply connected region, whose boundary consists (say) of a finite number of arcs of analytic curves.

In comparing this definition with the one given in §4, it will be observed that the exponent 2 has been omitted from the integrands in II, III, IV. However, Schwarz's inequality shows us that, if

$$\lim_{n=\infty} \iint_E |\Psi_n|^2 dy\, dt = 0,$$

then also

$$\lim_{n=\infty} \iint_E |\Psi_n|\, dy\, dt = 0,$$

provided that E is finite. In other words, a function $u(y, t)$ which satisfies the conditions of the former definition will surely satisfy the conditions of this last definition.

From the fundamental facts about convergence on the average, the reader will readily verify the truth of the following

LEMMA 1. *Let (y', t') be an interior point of E and δ a sufficiently small preassigned positive number. Then it is possible to find a subsequence, $u_n^*(y, t)$, of the sequence $u_n(y, t)$, such that for almost all choices of t_0 in the interval $|t'-t_0| \leqq \delta$ we shall have*

$$\lim_{k=\infty} \int_{E_{t_0}} \left| \frac{\partial u}{\partial t} - \frac{\partial u_k^*}{\partial t} \right| dy = 0,$$

where E_{t_0} denotes the cross-sectional point set obtained from E by putting $t=t_0$, and where $u(y, t)$ and $u_n(y, t)$ satisfy the requirements of Definition 1.

DEFINITION 2. *Let $u(y, t)$, defined on E, be a solution of $P(u)=0$ in the generalized sense, and let $u_n(y, t)$ be the approximation functions introduced in Definition 1. Then the cross-sectional point set E_{t_0} obtained from E by setting $t=t_0$ is called a proper line, if there exists a subsequence $u_n^*(y, t)$ such that*

$$\lim_{n=\infty} \int_{E_{t_0}} \left| \frac{\partial u}{\partial t} - \frac{\partial u_n^*}{\partial t} \right| dy = 0.$$

If such a subsequence does not exist, E_{t_0} is called an improper line.

According to Lemma 1, almost all cross-sections are proper lines.

DEFINITION 3. *A solution, $u(y, t)$, of $P(u)=0$ in the generalized sense will be said to assume the initial values $f(y)$ and $g(y)$ for $t=t_0$, if both the following conditions hold:*

(I) *E_{t_0} is a proper line;*

(II) *$u(y, t_0)=f(y)$, and*

$$\left.\frac{\partial u}{\partial t}\right|_{t=t_0} = g(y)$$

almost everywhere on E_{t_0}.

The subsequence $u_n^*(y, t)$ mentioned in Definition 2 will then be such that

$$\lim_{n=\infty} \int_{E_{t_0}} \left| \frac{\partial u_n^*}{\partial t} - g(y) \right| dy = 0.$$

We temporarily consider the simple equation

$$\frac{\partial^2 u}{\partial t^2} - \frac{\partial^2 u}{\partial y^2} = \phi(y, t). \tag{5.1}$$

LEMMA 2. *If $u(y, t)$ is a given function satisfying* (5.1) *in the ordinary sense, then*

$$\begin{aligned} u(y, t) &= \tfrac{1}{2}u(y + t - t_0, t_0) + \tfrac{1}{2}u(y - t + t_0, t_0) \\ &\quad + \frac{1}{2}\int_{y-t+t_0}^{y+t-t_0} \frac{\partial u}{\partial t}(\eta, t_0)d\eta + \frac{1}{2}\iint_{\mathrm{tri}(y,t,t_0)} \phi(\eta, \tau)d\eta d\tau. \end{aligned} \tag{5.2}$$

t_0 is arbitrary except as restricted below. The region of integration for the double integral on the right, denoted by tri (y, t, t_0), *is the triangle in the (η, τ) plane with vertices at the following points: (y, t), $(y-t+t_0, t_0)$, and $(y+t-t_0, t_0)$.*

We also require that tri(y, t, t_0) shall lie entirely within the region of definition of ϕ and u, and that ϕ shall be integrable, so that the right hand side of (5.2) will have a meaning.

This well known lemma can be easily proved by applying the following formula, deduced from Green's theorem, to tri (y, t, t_0):

$$\iint_S \left(\frac{\partial^2 u}{\partial y^2} - \frac{\partial^2 u}{\partial t^2}\right) d\eta d\tau = \int_C \frac{\partial u}{\partial t} d\eta + \frac{\partial u}{\partial y} d\tau,$$

where S represents any closed region in the (η, τ) plane and C represents the boundary of S taken in the proper sense. For later convenience I have written η and τ as the variables of integration instead of y and t. That is, in the above integrals, I regard $u(y, t)$ and its partial derivatives as being evaluated for $y = \eta$ and $t = \tau$. In the sequel, the notation will frequently be changed in this way, whenever no confusion is likely to result, with no further comment. The actual proof of the lemma is omitted.

From Lemma 2 we see that the non-linear partial differential equation (4.1) is closely related to the following non-linear integro-partial differential equation:

$$\begin{aligned} u(y, t) &= \tfrac{1}{2}u\,(y+t-t_0, t_0) + \tfrac{1}{2}u\,(y-t + t_0, t_0) \\ &\quad + \frac{1}{2}\int_{y-t+t_0}^{y+t-t_0} \frac{\partial u}{\partial t}\bigg|_{\tau=t_0} d\eta + \frac{1}{2}\iint_{\mathrm{tri}\,(y,t,t_0)} F\left[\frac{\partial u}{\partial y}, \frac{\partial u}{\partial t}, u(\eta, \tau), \eta, \tau\right] d\eta d\tau, \end{aligned} \tag{5.3}$$

which we now proceed to consider.

Definition 4. *A function $u(y, t)$, defined in E and admitting almost everywhere first partial derivatives, is a solution of* (5.3) *at the point (y', t') in E, if, for almost all values of t_0 (termed "proper" values) in the neighborhood of t',* (5.3) *is an identity in y and t in the neighborhood of (y', t'). It is a solution throughout E, if it is a solution at every interior point of E.*

Equations (4.1) and (5.3) are only partially equivalent because a solution of (4.1) in the ordinary sense must possess second derivatives, whereas a solution of (5.3) need not possess second derivatives. We shall see, however, that (4.1) and (5.3) are completely equivalent, if by a solution of (4.1) we mean a solution in the generalized sense, provided that F shall satisfy certain simple conditions.

Theorem 1. *Let $F(p_1, p_2, u, y, t)$ be defined for all p_1 and p_2, for $|u| \leqq h$, and for (y, t) in E. Let F also obey the Lipschitz condition*

$$|F(p, p, u, y, t) - F(\tilde{p}, \tilde{p}, \tilde{u}, y, t)| \leqq \alpha |p_1 - \tilde{p}_1| + \beta |p_2 - \tilde{p}_2| + \gamma |u - \tilde{u}|$$

for $|u| \leqq h$ and $|u| \leqq h$. Then a solution $u(y, t)$ of (4.1) *in the generalized sense, such that $|u(y, t)| < h$, is also a solution of* (5.3).

There exists a sequence of functions $u_n(y, t)$ satisfying the conditions of Definition 1. Because of Condition I we may assume without loss of generality that $|u_n| < h$, $n = 1, 2, 3, \cdots$.

Let (y', t') be any interior point of E. It is possible to choose a positive number δ and a neighborhood U of (y', t') such that, if (y, t) is a point of U and t_0 satisfies the inequality $|t' - t_0| \leqq \delta$, the triangular region tri(y, t, t_0) will lie completely imbedded in E.

In accordance with Lemma 1 we have for almost all choices of t_0 the following relation:

$$\lim_{n=\infty} \int_{E_{t_0}} \left| \frac{\partial u}{\partial t} - \frac{\partial u_n^*}{\partial t} \right| dy = 0, \tag{5.4}$$

where $u_n^*(y, t)$ is a certain subsequence of the given sequence, $u_n(y, t)$. In other words, $t = t_0$ is a proper line. Choose any such proper line such that $|t' - t_0| \leqq \delta$. Then hold t_0 fast.

If we set

$$\zeta_n(y, t) \equiv \frac{\partial^2 u_n^*}{\partial t^2} - \frac{\partial^2 u_n^*}{\partial y^2} - F\left(\frac{\partial u_n^*}{\partial y}, \frac{\partial u_n^*}{\partial t}, u_n^*, y, t\right),$$

we know from Definition 1, Condition IV, that

$$\lim_{n=\infty} \iint_E |\zeta_n(\eta, \tau)| d\eta d\tau = 0. \tag{5.5}$$

Also from Lemma 2 we have the following identity in y and t in the neighborhood U of (y', t'):

$$\begin{aligned} u_n^*(y, t) \equiv\ & \tfrac{1}{2}u_n^*(y + t - t_0, t_0) + \tfrac{1}{2}u_n^*(y - t + t_0, t_0) \\ & + \frac{1}{2}\int_{y-t+t_0}^{y+t-t_0} \frac{\partial u_n^*}{\partial t}(\eta, t_0)d\eta \\ & + \frac{1}{2}\iint_{\mathrm{tri}(y,t,t_0)} F\left[\frac{\partial u_n^*}{\partial y}, \frac{\partial u_n^*}{\partial t}, u^*, \eta, \tau\right]d\eta d\tau \\ & + \frac{1}{2}\iint_{\mathrm{tri}(y,t,t_0)} \zeta_n(\eta, \tau)d\eta d\tau. \end{aligned} \tag{5.6}$$

Now consider the function

$$\begin{aligned} w(y, t) \equiv\ & \tfrac{1}{2}u(y + t - t_0, t_0) + \tfrac{1}{2}u(y - t + t_0, t_0) \\ & + \frac{1}{2}\int_{y-t+t_0}^{y+t-t_0} \frac{\partial u}{\partial t}(\eta, t_0)d\eta + \frac{1}{2}\iint_{\mathrm{tri}(y,t,t_0)} F\left[\frac{\partial u}{\partial y}, \frac{\partial u}{\partial t}, u, \eta, \tau\right]d\eta d\tau. \end{aligned} \tag{5.7}$$

The fact that the double integral on the right actually exists follows from the fact that

$$\iint_E F\left[\frac{\partial u_n}{\partial y}, \frac{\partial u_n}{\partial t}, u_n, \eta, \tau\right]d\eta d\tau$$

is *assumed* to exist, since u was by hypothesis a solution in the generalized sense. It is easily proved in virtue of the Lipschitz condition and Conditions I, II, and III of Definition 1 that $F(\partial u_n/\partial y, \partial u_n/\partial t, u_n(y, t), y, t)$ converges in the mean, as n tends to infinity, to $F(\partial u/\partial y, \partial u/\partial t, u(y, t), y, t)$, and hence this latter function is integrable.

We shall show that $w(y, t) \equiv u(y, t)$ in the neighborhood U of (y', t'). Subtracting (5.6) from (5.7) we get

$$\begin{aligned} w(y,t) - u_n^*(y,t) =\ & \tfrac{1}{2}\{u(y+t-t_0,t_0) - u_n^*(y+t-t_0,t_0)\} + \tfrac{1}{2}\{u(y-t+t_0,t_0) \\ & - u_n^*(y - t + t_0, t_0)\} \\ & + \frac{1}{2}\int_{y-t+t_0}^{y+t-t_0} \left\{\frac{\partial u}{\partial t}(\eta, t_0) - \frac{\partial u_n^*}{\partial t}(\eta, t_0)\right\}d\eta \\ & + \frac{1}{2}\iint_{\mathrm{tri}(y,t,t_0)} \left\{F\left(\frac{\partial u}{\partial y}, \cdots\right) - F\left(\frac{\partial u_n^*}{\partial y}, \cdots\right)\right\}d\eta d\tau \\ & - \frac{1}{2}\iint_{\mathrm{tri}(y,t,t_0)} \zeta_n(\eta,\tau)\, d\eta d\tau. \end{aligned}$$

Let ϵ be an arbitrarily small positive number. Then the sum of the absolute values of the first two terms on the right can be taken less than $\frac{1}{4}\epsilon$ by taking n sufficiently large (independently of y or t) because of Condition I of Definition 1. The absolute value of the third term can be taken less than $\frac{1}{4}\epsilon$ because of (5.4). The absolute value of the fifth (last) term can also be taken less than $\frac{1}{4}\epsilon$ by (5.5). And finally applying the Lipschitz condition to the fourth term we have

$$
\begin{aligned}
| w(y,t) - u_n^*(y,t) | < \frac{3}{4}\epsilon &+ \frac{1}{2}\alpha\cdot \iint_{\mathrm{tri}(y,t,t_0)} \left| \frac{\partial u}{\partial y} - \frac{\partial u_n^*}{\partial y} \right| d\eta d\tau \\
&+ \frac{1}{2}\beta\cdot \iint_{\mathrm{tri}(y,t,t_0)} \left| \frac{\partial u}{\partial t} - \frac{\partial u_n^*}{\partial t} \right| d\eta d\tau \\
&+ \frac{1}{2}\gamma\cdot \iint_{\mathrm{tri}(y,t,t_0)} | u(\eta,\tau) - u_n^*\eta,\tau) | \, d\eta d\tau.
\end{aligned}
$$

Hence from Conditions I, II, III, we have $|w(y,t)-u_n^*(y,t)| < \epsilon$, if $n > N'$ where N is a number depending only upon ϵ. In other words $u_n^*(y,t)$ tends uniformly to $w(y,t)$ in U. Since, however, u_n^* is a subsequence of u_n, which by hypothesis converges uniformly to $u(y,t)$, it follows that $w(y,t) \equiv u(y,t)$ in the neighborhood of (y', t').

Hence u is a solution of (5.3) at the point (y', t'). Since (y', t') was any interior point of E, u is by Definition 4 a solution throughout E of (5.3).

THEOREM 2. *Let y', t', t_1 be any three real numbers determining a closed triangular region,* $\mathrm{tri}(y', t', t_1)$, *and such that $t'-t_1 = T > 0$. Let $F(p_1, p_2, u, y, t)$ be defined for all values of y and t which are coordinates of points in* $\mathrm{tri}(y', t', t_1)$; *for $|u| < h$, and for all values of p_1 and p_2 whatever. And suppose that it satisfies the Lipschitz condition*

$$
|F(\tilde{p}_1, \tilde{p}_2, \tilde{u}, y, t) - F(p_1, p_2, u, y, t)| \leqq \alpha|\tilde{p}_1 - p_1| + \beta|\tilde{p}_2 - p_2| + \gamma|\tilde{u} - u|,
$$

for $|u| < h$, $|\tilde{u}| < h$.

Let $f(y)$ and $g(y)$ be defined for $y' - T \leqq y \leqq y' + T$. Let $f(y)$ be an indefinite integral of a function $f'(y)$, and let $g(y)$ be summable.

Then there can not be more than one solution, $u(y,t)$, defined on $\mathrm{tri}(y', t', t_1)$, *of the equation*

$$
\begin{aligned}
u(y,t) &= \tfrac{1}{2}f(y+t-t_1) + \tfrac{1}{2}f(y-t+t_1) \\
&+ \frac{1}{2}\int_{y-t+t_1}^{y+t-t_1} g(\eta)d\eta + \frac{1}{2}\iint_{\mathrm{tri}(y,t,t_1)} F\left[\frac{\partial u}{\partial y}, \frac{\partial u}{\partial t}, u, \eta, \tau\right] d\eta d\tau,
\end{aligned}
\tag{5.8}
$$

for which $|u| < h$.

Let $u(y, t)$ and $v(y, t)$ be two such solutions of (5.8).

Differentiating, using the Lipschitz condition, and executing other obvious operations, we establish

$$\left.\begin{array}{l} 2\cdot|u(y,t)-v(y,t)| \\ \displaystyle\int_{y'-t'+t}^{y'+t'-t}\left|\frac{\partial u}{\partial t}(\eta,t)-\frac{\partial v}{\partial t}\right|d\eta \\ \displaystyle\int_{y'-t'+t}^{y'+t'-t}\left|\frac{\partial u}{\partial y}(\eta,t)-\frac{\partial v}{\partial y}\right|d\eta \end{array}\right\} \leqq \int_{t_1}^{t}\int_{y'-t'+\tau}^{y'+t'-\tau}\left\{\alpha\cdot\left|\frac{\partial u}{\partial y}-\frac{\partial v}{\partial y}\right|+\beta\cdot\left|\frac{\partial u}{\partial t}-\frac{\partial v}{\partial t}\right| + \gamma\cdot|u-v|\right\}d\eta d\tau.$$

The last two of the above three inequalities hold for almost all values of t on $t_1 \leqq t \leqq t'$. Let t_2 be one of these non-exceptional values, not greater than $t_1+\frac{1}{2}/(\alpha+\beta+T\gamma)$. Let M be the least common upper bound of the left members of the above inequalities for the non-exceptional values of t on the interval $t_1 \leqq t \leqq t_2$ and for y restricted so that the point (y, t) lies in tri(y', t', t_1). Hence, for some such point (y^*, t^*), one at least of these left members is not less than $\frac{3}{4}M$, and we thus obtain

$$\begin{aligned} \tfrac{3}{4}M &\leqq M(\alpha+\beta)(t_2-t_1) + 2T(t_2-t_1)\gamma M/2 \\ &= (t_2-t_1)(\alpha+\beta+T\gamma)M \leqq \tfrac{1}{2}M. \end{aligned}$$

Therefore $M=0$.

It follows that $u(y, t) \equiv v(y, t)$ in the part of tri(y', t', t_1) which lies below the line $t=t_2$. By a repetition of this argument the reader can readily extend this result to include the whole of tri (y', t', t_1).

Evidently Theorems 1 and 2 can be used to prove the uniqueness of the function $u(y, t)$ of §4, which satisfies (in the generalized sense) the differential equation (4.1) and obeys the conditions (4.3).

In order to apply Theorem 2, however, it is first necessary to extend the definitions of $F(p_1, p_2, u, y, t)$, $f(y)$, $g(y)$, and $u(y, t)$, which in §4 were regarded as defined only for $0 \leqq y \leqq \pi$. It is necessary to do this so that every point in the rectangle $0 \leqq y \leqq \pi$, $0 \leqq t \leqq K$ may be imbedded within a triangular region tri$(y', t', 0)$. This extension can obviously be effected in a variety of ways. For example, we could define F, f, g, u outside of $0 \leqq y \leqq \pi$ by making them periodic in y with period π. This definition alone may give conflicting values for F at points for which $y=k\pi$, $k=0, \pm 1, \pm 2, \pm 3, \cdots$. So we shall redefine F at these points by writing $F(p_1, p_2, u, k\pi, t) \equiv 0$. With this extended definition, $u(y, t)$ furnishes us with a generalized solution of (4.1) which is valid for $-N\pi \leqq y \leqq +N\pi$, $0 \leqq t \leqq K$, where N is an arbitrarily large integer. In order to see this it is only necessary *first* to extend the definitions of the approximation functions $u_n(y, t)$ by making them periodic in

y and *secondly* to modify them slightly near the points for which $y=k\pi$, so that they may be of class C'' throughout in accordance with Definition 1.

PART III

6. Statement of the problem for the parabolic partial differential equation. We now treat certain partial differential equations of parabolic type by the methods developed in Parts I and II. A different point of view is assumed, however, in that no use is made of a solution in a generalized sense. M. R. Siddiqi†, using the methods of Lichtenstein, has treated parabolic equations of a more restricted type and for less general initial conditions. The present methods are also simpler than Siddiqi's; but on the other hand Siddiqi's solution is valid for $0 \leqq t < \infty$, whereas the present solution is defined only for a sufficiently small interval $0 \leqq t \leqq K > 0$. The inequalities upon which the present work is based seem to admit considerable latitude, but I have been unable to obtain the extension to the infinite interval.

Since practically no repetition is involved, this part of the paper is written so that it can be read independently of Parts I and II.

We shall consider a partial differential equation of the form,

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial y^2} + F\left(\frac{\partial u}{\partial y}, u, y, t\right), \tag{6.1}$$

where $F(p, u, y, t)$ is defined for $|p| \leqq P$, $|u| \leqq U$, $0 \leqq y \leqq \pi$, $0 \leqq t \leqq T$. It is continuous and possesses continuous partial derivatives with respect to p, u, and y. Furthermore we assume either that

$$F(p, 0, y, t) \equiv 0 \tag{6.2}$$

or that

$$F(p, u, 0, t) \equiv F(p, u, \pi, t) \equiv 0.‡ \tag{6.2'}$$

As a consequence of the existence of the continuous partial derivatives we may also write the following Lipschitz condition:

$$|F(p, u, y, t) - F(\bar{p}, \bar{u}, y, t)| \leqq \alpha \cdot |p - \bar{p}| + \beta \cdot |u - \bar{u}|, \tag{6.3}$$

which is valid for the domain of definition specified above.

We have given a function $f(y)$, defined on $0 \leqq y \leqq \pi$, vanishing at the end

† See bibliography given in the Introduction.

‡ In the equations treated by Siddiqi, F is developable in a power series in u and p:

$$F(p, u, y, t) = \sum_{\alpha, \beta}^{0,1,\cdots,\infty} g_{\alpha\beta}(y, t) u^\alpha p^\beta,$$

where, however, α is not allowed to take on the value 0. Consequently Siddiqi's equations satisfy (6.2).

points of this interval, possessing an absolutely continuous first derivative, and having almost everywhere on $0 \leqq y \leqq \pi$ a second derivative whose square is summable. This is equivalent to the assumption that $f(y)$ can be developed in a trigonometric series $f(y) = \sum_{k=1}^{\infty} a_k \sin ky$, for which $\sum a_k^2 k^4$ converges.† We wish to find a function $u(y, t)$, defined for $0 \leqq y \leqq \pi$, $0 \leqq t \leqq K$, which satisfies (6.1) and the boundary conditions

$$u(y, 0) \equiv f(y),\ u(0, t) \equiv u(\pi, t) \equiv 0 \qquad (0 < K \leqq T). \tag{6.4}$$

7. The related infinite systems of ordinary differential equations. We shall here prove an existence lemma for infinite systems of differential equations of the type

$$\frac{dx_k}{dt} + \mu_k^2 x_k = f_k\{t, x\} \qquad (k = 1, 2, 3, \cdots) \tag{7.1}$$

under the initial conditions $x_k(0) = a_k$. Here, contrary to the notation of §1, a letter followed by $\{t, x\}$ denotes a function dependent upon the infinitely many independent variables $t, x_1, x_2, x_3, \cdots$. The μ_k are any infinite set of real numbers, such that $1 \leqq \mu_k \leqq \mu_{k+1}$. We shall consider (7.1) in the form of the equivalent infinite system of integral equations:

$$x_k(t) - a_k e^{-\mu_k^2 t} = \int_0^t f_k\{\tau, x(\tau)\} e^{\mu_k^2(\tau - t)} d\tau. \tag{7.2}$$

As in Part I, we shall use the following terminology and abbreviations: $[x]$ stands for the infinite system of numbers represented by the symbols $x_1, x_2, x_3, \cdots$. $\|x\|$ is an abbreviation for $(\sum_{k=1}^{\infty} x_k^2)^{1/2}$, if this limit exists. It will be noted that this symbol obeys the inequality

$$\|x + x'\| \leqq \|x\| + \|x'\|. \tag{7.3}$$

The ordered sequence $[x]$ will frequently be regarded as a point in function space. The region $Q(q)$ will consist of those points $[x]$ for which

$$\|\mu^2 x\| \leqq q, \tag{7.4}$$

where q is positive. The region $R(r)$ consists of those points $[x]$ for which

$$\|\mu^2(x - ae^{-\mu^2 t})\| \leqq r, \tag{7.5}$$

for at least one value of $t > 0$ $(r > 0)$.

We shall always assume that $\|\mu^2 a\|$ exists and we shall take $q = r + \|\mu^2 a\|$. Evidently, then, $R < Q$, as follows from (7.3).

† Siddiqi assumes for one of his theorems that $\sum k^2 |a_k|$ converges and for the other theorem that $\sum k^3 |a_k|$ converges.

We assume that $f_k\{t, x\}$ is defined for $0 \leqq t \leqq T$ and for $[x]$ in some region $Q(q)$. Furthermore we assume the following three fundamental hypotheses:

Hypothesis 1. *If $[x(t)]$ is any sequence of continuous functions whose range is in $Q(q)$ for $0 \leqq t \leqq T$, then $f_k\{t, x(t)\}$ is integrable on $0 \leqq t \leqq T$.*

Hypothesis 2. *A number B exists such that $\|\mu f\{t, x\}\| \leqq 2^{1/2} B$ for $[x]$ in $Q(q)$ and for $0 \leqq t \leqq T$.*

Hypothesis 3. *A number A exists such that $\|f\{t, x\} - f\{t, \bar{x}\}\| \leqq 2^{1/2} A \cdot \|\mu(x - \bar{x})\|$ for $0 \leqq t \leqq T$ and for $[x]$ and $[\bar{x}]$ in $Q(q)$.*

On the basis of these hypotheses *we shall prove the existence of an infinite sequence of continuous functions $[x(t)]$ defined for $0 \leqq t \leqq K$ (K is the smaller of the two numbers T and r^2/B^2) whose range is in $R(r)$, such that when these functions are substituted in* (7.2) *the right hand members exist and are identically equal to the left.* Such a set of functions is called a solution of (7.2). *This solution is unique.*

From Schwarz's inequality and Lebesgue's theorem for integrating infinite sequences, it is readily verified that, if $F_k(t)$ is a sequence of integrable functions such that $\sum_{k=1}^{n} \mu_k^2 [F_k(t)]^2$ is uniformly bounded with respect to n and t (for $0 \leqq t \leqq T$), then

$$\left\| \mu^{p+1} \int_0^t F(\tau) e^{\mu^2(\tau - t)} d\tau \right\|^2 \leqq \frac{1}{2} \int_0^t \|\mu^p F(\tau)\|^2 d\tau, \tag{7.6}$$

where $p = 0$ or 1 and $0 \leqq t \leqq T$.

We set up the successive approximations

$$x_k^{(0)}(t) = a_k e^{-\mu_k^2 t}, \cdots,$$

$$x_k^{(n)}(t) = a_k e^{-\mu_k^2 t} + \int_0^t f_k\{\tau, x^{(n-1)}(\tau)\} e^{\mu_k^2(\tau - t)} d\tau, \cdots.$$

Obviously $[x^{(0)}(t)]$ is in $R(r)$. Assume for the moment that $[x^{(n-1)}(t)]$ is in $R(r)$ for $0 \leqq t \leqq K$; then, if the $x_k^{(n-1)}(t)$ are also continuous, we have

$$\begin{aligned} \|\mu^2[x^{(n)}(t) - ae^{-\mu^2 t}]\|^2 &= \left\| \mu^2 \int_0^t f\{\tau, x^{(n-1)}(\tau)\} e^{\mu^2(\tau - t)} d\tau \right\|^2 \\ &\leqq \frac{1}{2} \int_0^t \|\mu f\{\tau, x^{(n-1)}(\tau)\}\|^2 d\tau \\ &\leqq B^2 t \leqq r^2 \text{ for } 0 \leqq t \leqq T, r^2/B^2. \end{aligned}$$

Hence we have shown by induction that $[x^{(n)}(t)]$ lies in $R(r)$ for $0 \leqq t \leqq K$, and that the $x_k^{(n)}(t)$ are continuous.

Also

$$\|\mu[x^{(1)}(t) - x^{(0)}(t)]\|^2 = \|\mu[x^{(1)}(t) - ae^{-\mu^2 t}]\|^2$$
$$\leqq \|\mu^2[x^{(1)}(t) - ae^{-\mu^2 t}]\|^2 \leqq B^2 t.$$

Using (7.6) and Hypothesis 3, we get

$$\|\mu[x^{(n+1)}(t) - x^{(n)}(t)]\|^2 = \left\|\mu\int_0^t [f\{\tau, x^{(n)}(\tau)\} - f\{\tau, x^{(n-1)}(\tau)\}]e^{\mu^2(\tau-t)}d\tau\right\|^2$$
$$\leqq \frac{1}{2}\int_0^t \|f\{\tau, x^{(n)}(\tau)\} - f\{\tau, x^{(n-1)}(\tau)\}\|^2 d\tau$$
$$\leqq \int_0^t A^2\cdot\|\mu[x^{(n)}(\tau) - x^{(n-1)}(\tau)]\|^2 d\tau.$$

We thus have a recursion formula which enables us to prove by induction that

$$\|\mu[x^{(n)}(t) - x^{(n-1)}(t)]\|^2 \leqq \frac{A^{2(n-1)}B^2 t^n}{n!}.$$

By the Weierstrass test we see that the successive approximations converge uniformly: $\lim_{n=\infty} x_k^{(n)}(t) = x_k(t)$ uniformly, this relation being regarded as defining the $x_k(t)$ for $0 \leqq t \leqq K$.

Using (7.3) we have

$$\|\mu[x(t) - x^{(n)}(t)]\| \leqq \sum_{m=n}^{\infty}\|\mu[x^{(m+1)}(t) - x^{(m)}(t)]\| \leqq \sum_{m=n}^{\infty}\left(\frac{A^{2m}B^2K^{m+1}}{(m+1)!}\right)^{1/2}.$$

Then from Hypothesis 3, we have

$$\|f\{t, x(t)\} - f\{t, x^{(n)}(t)\}\| \leqq 2^{1/2}A\sum_{m=n}^{\infty}\left(\frac{A^{2m}B^2K^{m+1}}{(m+1)!}\right)^{1/2}.$$

And, since the right hand member of this inequality may be taken arbitrarily small independently of $t(0 \leqq t \leqq K)$, we have $\lim_{n=\infty} f_k\{t, x^{(n)}(t)\} = f_k\{t, x(t)\}$ uniformly. It is therefore obvious that $[x(t)]$ constitutes a solution of (7.2). It also satisfies (7.1) almost everywhere on $0 \leqq t \leqq K$.

The proof of the uniqueness of this solution follows essentially the same lines and is left to the reader.

8. Application to the partial differential equation. In applying the results of the preceding article we shall take $\mu_k = k$; q = the lesser of the two numbers $6^{1/2}U/\pi$ and $6^{1/2}P/\pi$. And we shall assume (as in §7) that $q - \|a\mu^2\| = r > 0$. We define the $f_k\{t, x\}$ as follows.

Let $[x]$ be an arbitrary point of $Q(q)$. $\|\mu^2 x\| \leqq q$. Then from Schwarz's inequality we have

$$\sum_{k=1}^{\infty} k\,|x_k| = \sum_{k=1}^{\infty}(k^2\cdot|x_k|)\left(\frac{1}{k}\right) \leqq \left(\sum_{k=1}^{\infty} k^4 x_k^2\right)^{1/2}\left(\sum_{k=1}^{\infty}\frac{1}{k^2}\right)^{1/2}$$

$$= \|\mu^2 x\|\cdot\frac{\pi}{6^{1/2}} \leqq \frac{\pi q}{6^{1/2}} \leqq U \text{ and } P.$$

It follows that the two series $u'(y)=\sum_{k=1}^{\infty} kx_k \cos ky$ and $u(y)=\sum_{k=1}^{\infty} x_k \sin ky$ converge uniformly. Moreover $|u'(y)| \leqq P$ and $|u(y)| \leqq U$. By the Riesz-Fischer theorem we also have a function $u''(y)$, defined almost everywhere on $0 \leqq y \leqq \pi$, such that its square is summable on this interval and such that

$$(8.1) \qquad (2/\pi)\int_0^{\pi} u''(y) \sin ky\, dy = -k^2 x_k$$

and

$$(8.2) \qquad (2/\pi)\int_0^{\pi} [u''(y)]^2 dy = \|\mu^2 x\|^2 \leqq q^2.$$

Integrating (8.1) by parts, it is easily seen that $u''(y)$ is really the second derivative of $u(y)$, a fact anticipated in the notation. Thus corresponding to any point $[x]$ in $Q(q)$ we can write down a function $u(y)$ with the properties enunciated above. We are now in a position to use this function to define

$$(8.3) \qquad f_k\{t, x\} = (2/\pi)\int_0^{\pi} F[u'(y), u(y), y, t] \sin ky\, dy.$$

Integrating this by parts we have

$$f_k\{t, x\} = -\frac{2}{\pi}F[u'(y), u(y), y, t]\frac{\cos ky}{k}\Bigg|_0^{\pi} + \frac{2}{\pi}\int_0^{\pi}\frac{dF}{dy}\,\frac{\cos ky}{k}dy.$$

Hence from (6.2) or (6.2′) we get

$$kf_k\{t, x\} = \frac{2}{\pi}\int_0^{\pi}\left[\frac{\partial F}{\partial p}u''(y) + \frac{\partial F}{\partial u}u'(y) + \frac{\partial F}{\partial y}\right]\cos ky\, dy.$$

Therefore

$$\|\mu f\{t, x\}\|^2 = \frac{2}{\pi}\int_0^{\pi}\left[\frac{\partial F}{\partial p}u''(y) + \frac{\partial F}{\partial u}u'(y) + \frac{\partial F}{\partial y}\right]^2 dy.$$

Now the right hand side of this equality is bounded (with respect to $[x]$) because of two facts: I. The partial derivatives of F, being continuous in a closed region, are bounded. II. By Schwarz's inequality and (8.2) we know that any expression of the form $\int_0^{\pi} G(y)\, u''(y)\, dy$ is bounded whenever $G(y)$

is bounded. Hence Hypothesis 2 of the preceding article holds for the definition of $f_k\{t, x\}$ given in (8.3).

Now let $[x]$ and $[\bar{x}]$ be two points of Q, to which correspond respectively the two functions $u(y)$ and $\bar{u}(y)$. We have from (6.3)

$$\begin{aligned}
\|f\{t, x\} - f\{t, \bar{x}\}\|^2 &= \frac{2}{\pi}\int_0^\pi | F[u'(y), u(y), y, t] - F[\bar{u}'(y), \bar{u}(y), y, t] |^2 dy \\
&\leqq \frac{2}{\pi}\int_0^\pi [2\alpha^2 \cdot | u'(y) - \bar{u}'(y) |^2 + 2\beta^2 \cdot | u(y) - \bar{u}(y) |^2] dy \\
&= 2\alpha^2 \sum_{k=1}^\infty (kx_k - k\bar{x}_k)^2 + 2\beta^2 \sum_{k=1}^\infty (x_k - \bar{x}_k)^2 \\
&\leqq 2(\alpha^2 + \beta^2)\cdot \|\mu(x - \bar{x})\|^2.
\end{aligned}$$

Hence Hypothesis 3 holds with $A = (\alpha^2+\beta^2)^{1/2}$.

The proof that Hypothesis 1 holds is easy and is left to the reader.

Hence all the results of §7 are now available. Using the $x_k(t)$, the existence of which was asserted in that article, we form the function $u(y, t) = \sum_{k=1}^\infty x_k(t) \sin ky$. *Then $u(y, t)$ satisfies* (6.1) *almost everywhere in the region* $0 \leqq y \leqq \pi$, $0 \leqq t \leqq K$, *and also fulfills the boundary conditions* (6.4). *Furthermore $u(y, t)$ is the only function of class C' in y and possessing derivatives $\partial u/\partial t$, $\partial^2 u/\partial y^2$ almost everywhere, such that $|u(y, t)| \leqq U$, $|\partial u/\partial y| \leqq P$, which enjoys these properties.*

In the first place by (7.1), (7.4), and Hypothesis 2, it is seen that $\sum_{k=1}^\infty [dx_k/dt]^2$ converges. By the Riesz-Fischer theorem we can therefore find a function $w(y, t)$ whose Fourier coefficients are precisely these dx_k/dt. It is now an elementary exercise in analysis to identify $w(y, t)$ with $\partial u/\partial t$.

The theorem follows from the fact that the Fourier coefficients of

$$\frac{\partial u}{\partial t} - \frac{\partial^2 u}{\partial y^2} - F\left(\frac{\partial u}{\partial y}, u, y, t\right)$$

are none other than the

$$\frac{dx_k}{dt} + k^2 x_k - f_k\{t, x\},$$

all of which vanish. The uniqueness part of the theorem follows from the fact that the assumption of the existence of a second such function, $\bar{u}(y, t)$, leads to a contradiction of the uniqueness of the functions $[x(t)]$.

CAMBRIDGE, MASS.

ON DEFINITIONS OF BOUNDED VARIATION FOR FUNCTIONS OF TWO VARIABLES*

BY

JAMES A. CLARKSON AND C. RAYMOND ADAMS

1. **Introduction.** Several definitions have been given of conditions under which a function of two or more independent variables shall be said to be of bounded variation. Of these definitions six are usually associated with the names of Vitali, Hardy, Arzelà, Pierpont, Fréchet, and Tonelli respectively. A seventh has been formulated by Hahn and attributed by him to Pierpont; it does not seem obvious to us that these two definitions are equivalent, and we shall give a proof of that fact.

The relations between these several definitions have thus far been very incompletely determined, and there would appear to have been misconceptions concerning them. In the present paper we propose to investigate these relations rather fully, confining our attention to functions of two independent variables.

We first (§2) give the seven definitions mentioned above and a list of the known relations among them. In §3 some properties of the classes of functions satisfying the several definitions are established. In §4 we determine, for each pair of classes, whether one includes the other or they overlap. In §5 further relations are found concerning the extent of the common part of two or more classes. We next (§6) give a list of similar relations when only bounded functions are admitted to consideration; in §7 additional like relations are obtained when only continuous functions are admitted. We conclude (§8) with a list of the comparatively few relations that are not yet fully determined.

2. **Definitions.** The function $f(x, y)$ is assumed to be defined in a rectangle $R(a \leqq x \leqq b,\ c \leqq y \leqq d)$. By the term *net* we shall, unless otherwise specified, mean a set of parallels to the axes:

$$x = x_i\ (i = 0, 1, 2, \cdots, m),\quad a = x_0 < x_1 < \cdots < x_m = b;$$
$$y = y_j\ (j = 0, 1, 2, \cdots, n),\quad c = y_0 < y_1 < \cdots < y_n = d.$$

Each of the smaller rectangles into which R is divided by a net will be called a *cell*. We employ the notation

* Presented to the International Congress of Mathematicians, Zurich, September 5, 1932, and to the American Mathematical Society, April 15, 1933; received by the editors December 22, 1932.

$$\Delta_{11}f(x_i, y_j) = f(x_{i+1}, y_{j+1}) - f(x_{i+1}, y_j) - f(x_i, y_{j+1}) + f(x_i, y_j),$$
$$\Delta f(x_i, y_i) = f(x_{i+1}, y_{i+1}) - f(x_i, y_i).$$

The *total variation function*, $\phi(\bar{x})$ $[\psi(\bar{y})]$, is defined as the total variation of $f(\bar{x}, y)$ $[f(x, \bar{y})]$ considered as a function of $y[x]$ alone in the interval (c, d) $[(a, b)]$, or as $+\infty$ if $f(\bar{x}, y)$ $[f(x, \bar{y})]$ is of unbounded variation.

DEFINITION V (Vitali-Lebesgue-Fréchet-de la Vallée Poussin*). *The function $f(x, y)$ is said to be of bounded variation*† *if the sum*

$$\sum_{i=0,\, j=0}^{m-1,\, n-1} |\Delta_{11}f(x_i, y_j)|$$

is bounded for all nets.

DEFINITION F (Fréchet). *The function $f(x, y)$ is said to be of bounded variation if the sum*

$$\sum_{i=0,\, j=0}^{m-1,\, n-1} \epsilon_i \bar{\epsilon}_j \Delta_{11}f(x_i, y_j)$$

is bounded for all nets and for all possible choices of $\epsilon_i = \pm 1$ and $\bar{\epsilon}_j = \pm 1$.

DEFINITION H (Hardy-Krause). *The function $f(x, y)$ is said to be of bounded variation if it satisfies the condition of definition V and if in addition*‡ *$f(\bar{x}, y)$ is of bounded variation in y (i.e., $\phi(\bar{x})$ is finite) for at least one $\bar{x}$ and $f(x, \bar{y})$ is of bounded variation in x (i.e., $\psi(\bar{y})$ is finite) for at least one $\bar{y}$.*

DEFINITION A (Arzelà). *Let (x_i, y_i) $(i=0, 1, 2, \cdots, m)$ be any set of points satisfying the conditions*

$$a = x_0 \leqq x_1 \leqq x_2 \leqq \cdots \leqq x_m = b;$$
$$c = y_0 \leqq y_1 \leqq y_2 \leqq \cdots \leqq y_m = d.$$

Then $f(x, y)$ is said to be of bounded variation if the sum

$$\sum_{i=0}^{m-1} |\Delta f(x_i, y_i)|$$

is bounded for all such sets of points.

---

* References to most of the authors mentioned here in connection with the various definitions are given by Hahn, *Theorie der Reellen Funktionen*, Berlin, 1921, pp. 539–547, or by Hobson, *Theory of Functions of a Real Variable*, 3d edition, vol. 1, Cambridge, 1927, pp. 343–347. We need supplement these only by Tonelli, *Sulla quadratura delle superficie*, Accademia dei Lincei, Rendiconti, (6), vol. 3 (1926), pp. 357–362.

† *In the rectangle R* is always to be understood.

‡ The definition H as originally formulated imposed the two latter conditions for *every x* and *every y*, respectively, but it was shown by W. H. Young that the three conditions were redundant and that the definition could be reduced to the form given here. See Hobson, loc. cit., p. 345.

DEFINITION P (Pierpont). *Let any square net be employed, which covers the whole plane and has its lines parallel to the respective axes. The side of each square may be denoted by D, and no line of the net need coincide with a side of the rectangle R. A finite number of the cells of the net will then contain points of R, and we may denote by ω_ν the oscillation of $f(x, y)$ in the νth of these cells, regarded as a closed region. The function $f(x, y)$ is said to be of bounded variation if the sum*

$$\sum_{\nu} D\omega_\nu$$

is bounded for all such nets in which D is less than some fixed constant.

DEFINITION P_H (Hahn's version of definition P). *Let any net be employed in which we have $m=n$ and $x_{i+1}-x_i=(b-a)/m$, $y_{i+1}-y_i=(d-c)/m$ $(i=0, 1, 2, \cdots, m-1)$. Then there are m^2 congruent rectangular cells and we may let ω_ν' stand for the oscillation of $f(x, y)$ in the νth cell, regarded as a closed region. The function $f(x, y)$ is said to be of bounded variation if the sum*

$$\sum_{\nu=1}^{m^2} \frac{\omega_\nu'}{m}$$

is bounded for all m.*

DEFINITION T (Tonelli). *The function $f(x, y)$ is said to be of bounded variation if the total variation function $\phi(\bar{x})$ is finite almost everywhere in (a, b), and its Lebesgue integral over (a, b) exists (finite), while a symmetric condition is satisfied by $\psi(\bar{y})$.*

It may be of interest to indicate briefly how a set of definitions seemingly so diverse came to be formulated. Definition V, perhaps the most natural analogue of that of bounded variation for a function of one variable, is sufficient to insure the existence of the Riemann-Stieltjes double integral $\int_a^b\int_c^d g(x, y)d_xd_yf(x, y)$ for every continuous function $g(x, y)$. The existence of this integral†, when $g(x, y)$ is the product of a continuous function of x and a continuous function of y, is also implied by condition F, which is weaker than V (see relation (3) below). Definition H singles out a class of functions which it is convenient to consider in the study of double Fourier series. Definition A, also a rather natural analogue of that of bounded variation for a function

* It may readily be proved that this is equivalent to assuming that there exists some infinite sequence of values of m, say $m_k(k=1, 2, 3, \cdots ; m_{k-1}<m_k)$, with m_k/m_{k-1} bounded, for which this sum is bounded.

† In a certain restricted sense; see Fréchet, *Sur les fonctionnelles bilinéaires*, these Transactions, vol. 16 (1915), pp. 215–234, especially pp. 225–227. Several questions concerning these double integrals are considered in a forthcoming paper by Clarkson.

of one variable, expresses a condition necessary and sufficient that $f(x, y)$ be expressible as the difference of two bounded monotone functions.* Definition P, or P_H, is a natural extension to functions of two variables of the notion of *bounded fluctuation*, to use Hobson's terminology, which is equivalent to that of bounded variation for functions of one variable. Condition T is necessary and sufficient that the surface $z=f(x, y)$, where $f(x, y)$ is continuous, be of finite area in the sense of Lebesgue; this definition also is useful in connection with double Fourier series.

For simplicity we shall also use the letters V, F, H, A, P, P_H, and T to represent the classes of functions satisfying the respective definitions. The class of bounded functions will be denoted by B and the class of continuous functions by C; a product, such as $V\cdot T\cdot C$, will stand for the common part of the two or more classes named.†

The only relations that seem to be already known among the several definitions may be indicated as follows‡:

(1)	$P_H > A > H,$	(2)	$A\cdot C > H\cdot C,$	(3)	$F > V > H,$
(4)	$V\cdot C > H\cdot C,$	(5)	$T\cdot C > A\cdot C,$	(6)	$V\cdot T\cdot C = H\cdot C.$

3. **Some properties of functions belonging to these classes.**§ We first prove the following theorem.

THEOREM 1. *If $f(x, y)$ is in class H, the total variation function $\phi(\bar{x})$ $[\psi(\bar{y})]$ is of bounded variation in the interval (a, b) $[(c, d)]$.*‖

Assume the contrary; then, given any $M>0$, there exists a set of numbers x_i $(i=0, 1, 2, \cdots, n)$ with

* Monotone in the sense of Hobson, loc. cit., p. 343.

† From the definitions the following relations are easily seen: $V>V\cdot B$, $F>F\cdot B$, $T>T\cdot B$, $H=H\cdot B$, $A=A\cdot B$, $P=P\cdot B$, and $P_H=P_H\cdot B$.

‡ For a proof of the relation $A\geqq H$ see for example Hobson, loc. cit., pp. 345–346; the relation $A>H$ then follows from an example given by Küstermann, *Funktionen von beschränkter Schwankung in zwei reellen Veränderlichen*, Mathematische Annalen, vol. 77 (1916), pp. 474–481. Since Küstermann's example is continuous, it also gives us $A\cdot C>H\cdot C$. A proof of the relation $P_H>A$ is given by Hahn, loc. cit., pp. 546–547. From the definitions we clearly have $V\geqq H$, and the relations $V>H$ and $V\cdot C>H\cdot C$ may then be inferred from the example $f(x, y)=x \sin (1/x)(x\neq 0)$, $f(0, y)=0$. That F is $\geqq V$ is obvious from the definition; the definite inequality $F>V$ is established by Littlewood, *On bounded bilinear forms in an infinite number of variables*, Quarterly Journal of Mathematics, Oxford Series, vol. 1 (1930), pp. 164–174. The relations $T\cdot C>A\cdot C$ and $V\cdot T\cdot C=H\cdot C$ are stated by Tonelli, loc. cit.

§ Only properties of the total variation functions $\phi(\bar{x})$ and $\psi(\bar{y})$ are considered here; other properties will be examined in a forthcoming paper.

‖ This property is not enjoyed by all functions of class A; indeed it is easily seen (compare example (C) below) that $f(x, y)$ may be in A and yet ϕ and ψ be everywhere discontinuous. It is clear that if $f(x, y)$ is in V, $\phi[\psi]$ is either everywhere infinite or of bounded variation.

$$a = x_0 < x_1 < x_2 < \cdots < x_n = b$$

and such that

$$\sum_{i=1}^{n} | \phi(x_i) - \phi(x_{i-1}) | > M.$$

Consider any two successive points (x_{i-1}, c) and (x_i, c). From the definition of $\phi(\bar{x})$, there exists a set of points p_j on the line $x = x_i$, and their projections p_j' on the line $x = x_{i-1}$, such that we have

$$\left| \sum | f(p_j) - f(p_{j-1}) | - \sum | f(p_j') - f(p_{j-1}') | \right| \geqq | \phi(x_i) - \phi(x_{i-1}) | /2.$$

Hence for the net N composed of the boundary lines of the rectangle, the lines $x = x_{i-1}$ and $x = x_i$, and the horizontal lines through the points p_j, the V-sum is

$$V_N(f) = \sum_{i=0,\, j=0}^{m-1,\, n-1} | \Delta_{11} f(x_i, y_j) | \geqq | \phi(x_i) - \phi(x_{i-1}) | /2.$$

By a repetition of this process for each interval (x_{i-1}, x_i) we may prove the existence of a net N' for which the V-sum $V_{N'}(f)$ is $\geqq M/2$, thus contradicting the hypothesis that $f(x, y)$ is of class H.

Before proving Theorem 2 we demonstrate the following lemma, due essentially to Borel.†

Lemma 1. *Let E be a bounded set of positive interior measure and let the sequence of functions $\{f_n(x)\}$ defined on E converge to the limit function $f(x)$ at each point of E. Then, if ϵ is any positive number and if $E_n(\epsilon)$ denotes the subset of E where*

$$f(x) - f_n(x) | > \epsilon,$$

we have

$$\lim_{n \to \infty} m_i E_n(\epsilon) = 0.$$

Since $m_i E_n(\epsilon)$ is the least upper bound of the measures of the measurable subsets of $E_n(\epsilon)$, it suffices to show that if $\{E_n'(\epsilon)\}$ is any sequence of measurable sets contained respectively in $\{E_n(\epsilon)\}$, then $\lim_{n\to\infty} mE_n'(\epsilon)$ is 0. This will be true if E^*, the complete limit of $\{E_n'(\epsilon)\}$, is of measure zero; but E^* is a null set, since the sequence $\{f_n(x)\}$ cannot converge at any point of E^*.

† See Borel, *Leçons sur les Fonctions de Variables Réelles*, Paris, 1905, p. 37, where essentially this lemma is indicated but not proved.

THEOREM 2. *If a function $f(x, y)$ is in class P_H, and E is the set of points $\bar{x}[\bar{y}]$ in the interval (a, b) $[(c, d)]$ for which $\phi(\bar{x})$ $[\psi(\bar{y})]$ is infinite, then m_iE is zero.*†

In particular, if E is measurable (as it would be if for example $f(x, y)$ were continuous; cf. Theorem 4), it is of measure zero.

To prove Theorem 2, assume $f(x, y)$ is in P_H and that E, the set of points $\bar{x}$ for which $\phi(\bar{x})$ is infinite, is of positive interior measure. On E define the sequence of functions $f_n(x)$ as follows. For a fixed n, let R be divided by a net N into n^2 congruent rectangles, and at the point $\bar{x}$ of E let

$$g_n(\bar{x}) = \sum_{i=1}^{n}[\text{oscillation of } f(\bar{x}, y) \text{ in the interval } y_{i-1} \leqq y \leqq y_i].$$

Let $f_n(\bar{x}) = 1/g_n(\bar{x})$. Then $\lim_{n\to\infty} f_n(x)$ is 0 at each point of E, and hence at each point of E', some arbitrarily selected measurable subset of E of positive measure.

Let $\epsilon > 0$ be given. By Lemma 1 there exists an r such that $m_iE_r(\epsilon)$ is $<\epsilon$, where $E_r(\epsilon)$ is the subset of E' on which $|f_r(x)|$ is $>\epsilon$. Consider the net N which defines $f_r(x)$. Let λ be the number of columns of N in which points of the set $E' - E_r(\epsilon)$ occur. We have

$$\text{(a)} \qquad m_e[E' - E_r(\epsilon)] \leqq \lambda(b - a)/r.$$

From the relations

$$m_e[E' - E_r(\epsilon)] = mE' - m_iE_r(\epsilon) \quad \text{and} \quad m_iE_r(\epsilon) < \epsilon$$

we have

$$m_e[E' - E_r(\epsilon)] > mE' - \epsilon,$$

and hence by (a)

$$\lambda > r(mE' - \epsilon)/(b - a).$$

But in each of the λ columns of N which contain points of the set $E' - E_r(\epsilon)$ the sum of the oscillations of $f(x, y)$ in the several cells is at least $1/\epsilon$. Hence for the net N we have

$$\frac{1}{r}\sum_{\nu=1}^{r^2}\omega_\nu' \geqq \lambda/(\epsilon r) > (mE' - \epsilon)/[\epsilon(b - a)].$$

Since mE' is >0, this last quantity increases indefinitely with $1/\epsilon$, while if $f(x, y)$ is in P_H the sum on the left must be bounded.

---

† If $f(x, y)$ is in A, $\phi[\psi]$ is clearly bounded.

The part of the theorem concerning $\psi(\bar{y})$ may of course be demonstrated in the same manner.

We may note here that *the set E of points $\bar{x}$ for which $\phi(\bar{x})$ is infinite may nevertheless be everywhere dense in the interval* (a, b), as the following example shows. Define the function $f(x, y)$ on the unit square $I(0 \leqq x \leqq 1, 0 \leqq y \leqq 1)$ as follows. Let the rational points of the segment $0 \leqq x \leqq 1$ be enumerated and designated by $x_1, x_2, x_3, \cdots, x_n, \cdots$. On each line $x = x_j$ define $f(x, y)$ as 1 for y irrational and $> 1 - 1/2^j$, and zero otherwise. When x is irrational let $f(x, y)$ be zero for all y. For convenience we denote by S_j the segment $1 - 1/2^j \leqq y \leqq 1$ of the line $x = x_j$.

Clearly $f(x, y)$ is of unbounded variation in y for each fixed rational x, and these points are everywhere dense in the interval $(0, 1)$. But $f(x, y)$ is in P_H. For consider any square net of n^2 cells on I. In all cells of such a net except for those which contain more than one point of some segment S_j, the oscillation is zero; in the remainder the oscillation is 1. Let M be the number of the latter. Then M is at most equal to $M_1 + M_2 + M_3 + \cdots + M_p + n$, where M_j is the number of cells containing more than one point of S_j, and p is the largest integer for which $1/2^p$ exceeds $1/n$. But M_j is less than $2 + n/2^{j-1}$; hence we have

$$\sum_{\nu=1}^{n^2} \omega_\nu' = M/n < 5$$

and $f(x, y)$ is in P_H.

As a preliminary to the proof of our third theorem we shall first establish another lemma.

Let A denote any set of k real numbers,

$$A: \qquad a_1, a_2, a_3, \cdots, a_k,$$

and let $\theta = \sum_{i=1}^{k} |a_i|$. With this set we may associate 2^k sums of the form

$$\pm a_1 \pm a_2 \pm a_3 \pm \cdots \pm a_k.$$

These sums occur in 2^{k-1} pairs, of opposite sign, $\pm S_j$ $(j = 1, 2, 3, \cdots, 2^{k-1})$, the subscripts being assigned arbitrarily. Let S_j $(j = 1, 2, 3, \cdots, 2^{k-1})$ be that one of the jth pair which is positive, or zero if each sum in the pair vanishes. Denote by $\sum A$ the sum $\sum_{j=1}^{2^{k-1}} S_j$.

LEMMA 2. *We have $\sum A \geqq M_k \theta$, where*

$$M_k = \begin{cases} (k-1)! \Big/ \left[\left(\dfrac{k-1}{2}\right)!\right]^2 & \textit{for } k \textit{ odd}, \\ k!/(2[(k/2)!]^2) & \textit{for } k \textit{ even}. \end{cases}$$

Since we shall make use of this result only for k odd, and a similar proof can be given for k even, we confine ourselves to the

Proof for k odd. Without loss of generality we may assume the a_i to be non-negative, since both θ and $\sum A$ are invariant under the change of sign of any a_i.

In the particular case in which all the a_i are equal we have $\sum A = M_k\theta$. For let $[S_j]_h$ $(h=0, 1, 2, \cdots, (k-1)/2)$ denote in this case the set of expressions of the form $\pm\theta/k\pm\theta/k\pm\theta/k\pm \cdots \pm\theta/k$ in which exactly h minus signs occur. Then each S_j in $[S_j]_h$ has the value $(k-2h)\theta/k$. In $[S_j]_h$ there will be exactly $\binom{k}{h}$ sums S_j. Hence, adding, we obtain $\sum A = M_k\theta$.

We wish to show that *in every case* $\sum A \geqq M_k\theta$. Let A be any set and let a' and a'' be any two elements of A. Let S_j' $(j=1, 2, 3, \cdots, n)$ be the 2^{k-3} sums obtained from the set composed of the remaining elements of A. Then the 2^{k-1} sums S_j may be written in an array of four columns thus:

$$\begin{array}{llll}
S_1' + a' + a'', & |S_1' - a' - a''|, & |S_1' + a' - a''|, & |S_1' - a' + a''|, \\
S_2' + a' + a'', & |S_2' - a' - a''|, & |S_2' + a' - a''|, & |S_2' - a' + a''|, \\
\cdots & \cdots & \cdots & \cdots \\
S_n' + a' + a'', & |S_n' - a' - a''|, & |S_n' + a' - a''|, & |S_n' - a' + a''|.
\end{array}$$

If we denote by C_1, C_2, C_3, and C_4 the sums of the respective columns, we have $\sum A = C_1+C_2+C_3+C_4$. By comparison with the sum obtained when absolute value signs are omitted from the third and fourth columns, we have at once

$$\sum A \geqq C_1 + C_2 + 2\sum_{j=1}^{n} S_j'.$$

But if in the set A we replace a' and a'' each by $(a'+a'')/2$ to form the set A', we see that $\sum A'$ is precisely the right-hand member of this inequality. Therefore, *if in a set A any two elements are each replaced by their arithmetic mean, $\sum A$ is not increased.*

Now assume the existence of a set A of k elements with $\sum_{i=1}^{k} a_i = \theta$ (and hence with arithmetic mean θ/k), and with $\sum A = M_k\theta - \delta$, where δ is some positive number. Let ξ be the absolute value of the greatest deviation from θ/k of any one a_i. There are a finite number of the a_i, then, whose deviation in absolute value exceeds $\xi/2$. Select one such and pair with it some element whose algebraic deviation is of the opposite sign. Replace each of these by half their sum. By repeating this operation we form the set A_1, with the same arithmetic mean θ/k, for which the greatest deviation from the mean does not exceed $\xi/2$, while $\sum A_1$ is $\leqq \sum A$. This process may be repeated as many times as we may desire, to yield a set A_p whose elements deviate from

the mean by as little as we wish. But as $\sum A$ is evidently a continuous function of the elements of A, for sufficiently large p we must have both

$$\left| \sum A_p - M_k\theta \right| < \delta \text{ and } \sum A_p \leqq \sum A = M_k\theta - \delta.$$

From this contradiction follows Lemma 2 for k odd.

We may now prove

THEOREM 3. *If $f(x, y)$ is in class F, the total variation function $\phi(\bar{x})$ $[\psi(\bar{y})]$ is either everywhere infinite, or is bounded and integrable in the sense of Riemann* over the interval (a, b) $[(c, d)]$.*

Let $f(x, y)$ be in class F, and for some x_0 $(a \leqq x_0 \leqq b)$ let $\phi(x_0) = M_1$, a finite number. Consider the function $f'(x, y) = f(x, y) - f(b, y)$. Clearly $f'(x, y)$ is also in class F.

Let x_1 $(a \leqq x_1 \leqq b)$ be any number distinct from x_0, and let (x_1, y_i) $(i = 0, 1, 2, \cdots, n)$ be any set of $n+1$ points on the line $x = x_1$ with

$$c = y_0 < y_1 < y_2 < \cdots < y_n = d.$$

We have

$$f(x, y) = f(x_0, y) + f'(x, y) - f'(x_0, y),$$

whence

$$\begin{aligned}
&\sum_{i=1}^{n} \left| f(x_1, y_i) - f(x_1, y_{i-1}) \right| \\
&\quad = \sum_{i=1}^{n} \left| f(x_0, y_i) + f'(x_1, y_i) - f'(x_0, y_i) - f(x_0, y_{i-1}) - f'(x_1, y_{i-1}) \right. \\
&\qquad \left. + f'(x_0, y_{i-1}) \right| \\
&\quad \leqq \sum_{i=1}^{n} \left| f(x_0, y_i) - f(x_0, y_{i-1}) \right| + \sum_{i=1}^{n} \left| \Delta_{11} f'(x_0, y_{i-1}) \right| .
\end{aligned}$$

Since $f'(x, y)$ is in F, there exists a number M_2 such that we have

$$\sum_{i=0,\, j=0}^{m-1,\, n-1} \epsilon_i \bar{\epsilon} \Delta_{11} f'(x_i, y_j) < M_2$$

for any net. But $\sum_{i=1}^{n} \left| \Delta_{11} f'(x_0, y_{i-1}) \right|$ is the sum of the absolute values of the differences $\Delta_{11} f'(x_i, y_j)$ in one column of cells of the net composed of the

* It is easily seen from the proof that discontinuities of $\phi[\psi]$ can occur only at a denumerable set of points; indeed, for any $\epsilon > 0$, the number of points at which $\phi[\psi]$ has a saltus $> \epsilon$ is finite. It is appropriate to remark also that example (E) below shows that $\phi[\psi]$ may be bounded but not of bounded variation.

four vertical lines $x=a$, $x=x_0$, $x=x_1$, $x=b$ and the $n+1$ horizontal lines $y=y_i$ $(i=0, 1, 2, \cdots, n)$. Hence, a fortiori, it is less than M_2 and we have

$$\sum_{i=1}^{n} | f(x_1, y_i) - f(x_1, y_{i-1}) | < M_1 + M_2;$$

thus $\phi(\bar{x})$ is bounded by the latter number.

Now assume that $\phi(\bar{x})$ is bounded but not integrable in the Riemann sense. Then E, the set of points in the interval (a, b) at which $\phi(\bar{x})$ is discontinuous, must be of positive exterior measure. Let E_n be the subset of E such that at each point of E_n the saltus of $\phi(\bar{x})$ exceeds $2/n$. Then E is $\sum_{n=1}^{\infty} E_n$, and so for some fixed k we must have $m_e E_k > 0$. Let x_1 be any point of E_k and let Δ_1 be an interval of length not exceeding $\frac{1}{2} m_e E_k$, with center x_1. Within Δ_1 there must be a point x_1' such that $|\phi(x_1) - \phi(x_1')| > 1/k$. We may assume without loss of generality that $\phi(x_1)$ is $> \phi(x_1')$. Let m and M be any constants satisfying the inequalities

$$\phi(x_1') < m < M < \phi(x_1), \quad M - m > 1/k.$$

Then there exists a set of points on the line $x=x_1$,

$$p_0(x_1, c), p_1, p_2, \cdots, p_r(x_1, d),$$

and their horizontal projections q_i $(i=0, 1, 2, \cdots, r)$ on the line $x=x_1'$, such that we have

$$\sum_{i=1}^{r} | f(p_i) - f(p_{i-1}) | > M, \quad \sum_{i=1}^{r} | f(q_i) - f(q_{i-1}) | < m,$$

and hence

$$\text{(b)} \qquad \sum_{i=1}^{r} | f(p_i) - f(p_{i-1}) - f(q_i) + f(q_{i-1}) | > M - m > 1/k.$$

Now consider the net N_1 on R consisting of the four vertical lines $x=a$, $x=x_1$, $x=x_1'$, $x=b$, and the $r+1$ horizontal lines through the points p_i $(i=0, 1, 2, \cdots, r)$. From (b) it is seen that the sum of the absolute values of the terms $\Delta_{11} f(x_i, y_j)$ associated with the single column of cells which stands on the interval (x_1, x_1') exceeds $1/k$.

Since the length of Δ_1 is $\leqq m_e E_k/2$, there is a point x_2 of E_k exterior to Δ_1. Surround x_2 with an interval Δ_2 of which it is the center, of length not exceeding $m_e E_k/4$ and small enough so that it does not overlap Δ_1. Proceeding as before, we prove the existence of a second net N_2 of which one column of cells possesses the property that the sum of the absolute values of the terms $\Delta_{11} f(x_i, y_j)$ associated with it exceeds $1/k$. It is clear that the net composed of

all the lines in both N_1 and N_2 has two distinct columns of cells each possessing this property.

This process may be repeated indefinitely, and so we see that there exists a $\theta>0$ such that, given any integer k, there exists a net on R in at least k columns of which the sum of the absolute values of the differences $\Delta_{11}f(x_i, y_j)$ in the several cells of the column exceeds θ. We proceed to show that under these conditions the sum

$$F_N(f) = \sum_{i=0,\,j=0}^{m-1,\,n-1} \epsilon_i \bar{\epsilon}_j \Delta_{11} f(x_i, y_j)$$

may be made arbitrarily large by proper choice of the net N and the ϵ_i's and $\bar{\epsilon}_j$'s.

Let k (taken odd for convenience) be given, and let N be a net, of n rows and m columns of cells, such that in at least k columns the above condition is satisfied. Consider the matrix $\|a_{ij}\|$ for which $a_{ij}=\Delta_{11} f(x_{i-1}, y_{j-1})$ and in which all the a_{ij} but those arising from the k columns noted above are suppressed. This matrix has, then, n rows and k columns; renumbering the columns consecutively, we have

$$\|a_{ij}\| = \left\| \begin{matrix} a_{1n} & a_{2n} & \cdots & a_{kn} \\ \cdot & \cdot & \cdot & \cdot \\ a_{12} & a_{22} & \cdots & a_{k2} \\ a_{11} & a_{21} & \cdots & a_{k1} \end{matrix} \right\|,$$

with $\sum_{j=1}^{n} |a_{ij}|>\theta$ $(i=1, 2, 3, \cdots, k)$.

If it can be shown that the sum

$$F' = \sum_{i=1,\,j=1}^{k,\,n} \delta_i \bar{\delta}_j a_{ij} \qquad (|\delta_i| = |\bar{\delta}_j| = 1)$$

for some choice of the δ_i's and $\bar{\delta}_j$'s is arbitrarily large with k, the proof will be complete, since max $F_N(f)$ is $\geqq$ max F'.

The δ_i's may be chosen in 2^{k-1} essentially distinct ways. Let S_{jp} represent the absolute value of the sum yielded by the jth row of $\|a_{ij}\|$ with the pth such choice, and let

$$F_p' = \sum_{j=1}^{n} S_{jp} \qquad (p = 1, 2, 3, \cdots, 2^{k-1}).$$

Then each F_p' is a particular value of F' corresponding to some choice of the δ_i's and $\bar{\delta}_j$'s. We may write

$$\sum_{p=1}^{2^{k-1}} F_p' = \sum_{j=1}^{n} \left[\sum_{p=1}^{2^{k-1}} S_{jp} \right],$$

and by Lemma 2 we have

$$\sum_{p=1}^{2^{k-1}} S_{ip} \geqq M_k\theta_j \quad (j = 1, 2, 3, \cdots, n), \text{ where } \theta_j = \sum_{i=1}^{k} |a_{ij}|,$$

whence

$$\sum_{p=1}^{2^{k-1}} F_p' \geqq \sum_{j=1}^{n} M_k\theta_j \geqq M_k k\theta = \frac{k!}{\left[\left(\frac{k-1}{2}\right)!\right]^2}\theta,$$

since k is odd. *It follows that at least one F' exceeds*

$$\frac{k!}{\left[\left(\frac{k-1}{2}\right)!\right]^2\cdot 2^{k-1}}\theta.$$

By Stirling's formula we see that this quantity increases without limit with k; therefore the sum $F_N(f)$ can have no bound, contrary to the hypothesis that $f(x, y)$ is in class F. Thus Theorem 3 is proved.

THEOREM 4. *If $f(x, y)$ is in class C, then $\phi(\bar{x})$ and $\psi(\bar{y})$ are lower semi-continuous functions.**

Let the interval (c, d) be divided into 2^n equal parts by the numbers

$$y_0 = c, y_1, y_2, \cdots, y_{2^n} = d$$

and set

$$\phi_n(\bar{x}) = \sum_{i=1}^{2^n} | f(\bar{x}, y_i) - f(\bar{x}, y_{i-1}) | .$$

Since $f(\bar{x}, y)$ is continuous in y, we have $\lim_{n\to\infty} \phi_n(\bar{x}) = \phi(\bar{x})$, and since $f(\bar{x}, y_i)$ $(i=0, 1, 2, \cdots, n)$ is continuous in $\bar{x}$, $\phi_n(\bar{x})$ is a continuous function of $\bar{x}$. Moreover the sequence $\{\phi_n(\bar{x})\}$ is non-decreasing; hence $\phi(\bar{x})$ is lower semi-continuous.† Similarly $\psi(\bar{y})$ is of like character.

4. **Relations between pairs of classes.** We shall establish the following:

(7) $P = P_H$, (8) $T > H$, (9) $V \gtreqless P$, $P \not> V$,

(10) $A \gtreqless V$, $V \not> A$, (11) $V \gtreqless T$, $T \not> V$, (12) $A \gtreqless T$, $T \not> A$,

(13) $P \gtreqless T$, $T \not> P$, (14) $F > V$, (15) $F \gtreqless A$, $A \not> F$,

(16) $F \gtreqless P$, $P \not> F$, (17) $F \gtreqless T$, $T \not> F$.

---

* It is apparent from the proof that the hypothesis that $f(x, y)$ be continuous in each variable separately is sufficient here.

† See, for example, Hobson, loc. cit., 2d edition, vol. 2, 1926, p. 149.

Proof of (7). Let us first assume $f(x, y)$ to be in class P and consider *any* net of n^2 cells as in definition P_H. Without loss of generality we may suppose $d-c\geqq b-a$. Then there exists a net of square cells as used in definition P for which we have $D=(b-a)/n$ and whose vertical lines include $x=a$ and $x=b$. No square of the P net can overlap more than two cells of the P_H net; hence we have

$$\sum\omega_\nu'/n \leqq \frac{2}{b-a}\cdot \sum D\omega_\nu,$$

which is bounded. This establishes the relation $P\leqq P_H$.

Now assume $f(x, y)$ to be in P_H. Again let us suppose $d-c\geqq b-a$, and consider any square net N, as used in definition P, for which we have $D<(b-a)/2$. Let n be the largest integer satisfying the inequality $(b-a)/n \geqq D$; then we have

$$\frac{b-a}{n+1} < D \leqq \frac{b-a}{n} \quad \text{and} \quad \frac{b-a}{n} < 2\frac{b-a}{n+1} < 2D,$$

and therefore

$$\text{(c)} \qquad D \leqq (b-a)/n < 2D.$$

Consider now the net N' of n^2 cells as used in the P_H definition. From the second part of (c) it is seen that one cell of N' can overlap no more than three columns of cells of N. The height of one cell of N' is $(d-c)/n$, and if p is the smallest integer satisfying the inequality $p\geqq(d-c)/(b-a)$, we have

$$(d-c)/n \leqq p(b-a)/n < 2pD;$$

hence one cell of N' can overlap no more than $2p+1$ rows of cells of N. Thus we have

$$\sum D\omega_\nu \leqq 3(2p+1)\sum D\omega_\nu' \leqq 3(2p+1)(b-a)\sum \omega_\nu'/n,$$

which is bounded. This establishes the relation $P_H\leqq P$, and we conclude the identity of the two classes.

Proof of (8). It was shown in §3 that if $f(x, y)$ is of class H, then the total variation functions $\phi(\bar{x})$ and $\psi(\bar{y})$ are of bounded variation. From this follows at once $T\geqq H$. Then from the example*

$$\text{(A)} \qquad f(x, y) = \begin{Bmatrix} 0, & x<y \\ 1, & x\geqq y \end{Bmatrix} \text{ in I, the unit square,}$$

which is in T but not H, we infer (8).

* See Hahn, loc. cit., p. 547.

Proof of (9). The first part follows from example (A), which is in P but not V. Example

$$\text{(B)} \qquad f(x, y) = \left\{ \begin{array}{ll} x \sin(1/x), & x \neq 0 \\ 0, & x = 0 \end{array} \right\} \text{ in I,}$$

which is in V but not P, establishes the second part.

Proof of (10). The first of these relations follows from the second of (9). By taking sets of points (x_i, y_i) along the perimeter of the rectangle R, one sees immediately that a function of class A must satisfy the two latter conditions of definition H. Since there exist functions which are in A but not H, and by the last remark these must fail to be in V, the second of relations (10) follows.

Proof of (11). The first of these relations is shown by example (A), which is in T but not V, while example (B) shows the second.

Proof of (12). Example (A) establishes the first relation. The second is a consequence of the following example.

(C) Let E be a non-measurable set in the interval $0 \leqq x \leqq 1$, and let E' be the set of points on the downward sloping diagonal of I whose projection on the x-axis is E. Define $f(x, y)$ as 1 at all points of E' and zero at all other points of I. Then clearly $f(x, y)$ is in A; but it is not in T, since $\phi(\bar{x})$ is not measurable.

Proof of (13). The first relation follows from example

$$\text{(D)} \qquad f(x, y) = \left\{ \begin{array}{l} 1 \text{ for } x \text{ and } y \text{ both rational} \\ 0 \text{ otherwise} \end{array} \right\} \text{ in I,}$$

which is in T but not P. The second follows from the second of (12).

Proof of (14). It has already been remarked that this relation, which is included in (3), has been established by Littlewood. His proof, however, depends upon the theory of bilinear forms in infinitely many variables; it may therefore be of interest to show how an example of a function which is in class F but not V can be constructed directly. Moreover, we can easily determine whether our example belongs to the classes P, A, and T; consequently it may be expected to be useful in proving other relations later.

We first make a preliminary observation. Consider a function $f(x, y)$ defined in R. For any net N let max $F_N(f)$ denote the maximum value which the sum

$$F_N(f) = \sum_{i=0,\, j=0}^{m-1,\, n-1} \epsilon_i \bar{\epsilon}_j \Delta_{11} f(x_i, y_j)$$

associated with N may be made to assume by a suitable choice of the ϵ_i's

and $\bar{\epsilon}_j$'s. If an additional line, horizontal or vertical, be added to N to form the new net N', we have $\max F_N(f) \leqq \max F_{N'}(f)$. For, suppose a horizontal line be added. Then one row of cells of N is replaced by two new rows of cells of N'; and if the two $\bar{\epsilon}$'s associated with these rows in the sum $F_{N'}(f)$ be both assigned the same value as the $\bar{\epsilon}$ associated with the single replaced row in the sum $F_N(f)$, and all the remaining ϵ's and $\bar{\epsilon}$'s given identical values in the two sums, we have $F_N(f) = F_{N'}(f)$, from which the above observation follows.

By a "point-rectangle function" we shall mean a function $f(x, y)$ defined on R as follows: $f(x, y) = \pm 1$ (or some other constant) on each of a rectangular array of points p_{ij} in R, where the rows are equally spaced with each other and with the lines $y = c$, $y = d$, and the columns likewise, and p_{ij} is the point standing in the jth row and ith column of the array; $f(x, y) = 0$ at all other points of R. Let $\max F(f)$ denote the maximum value which the expression $F_N(f)$ can attain for all possible nets and choices of the ϵ's and $\bar{\epsilon}$'s, and $\max V(f)$ be the maximum value which the sum

$$V_N(f) = \sum_{i=0,\,j=0}^{m-1,\,n-1} \left| \Delta_{11} f(x_i, y_j) \right|$$

can attain with all possible nets N. We consider the problem of determining the value of $(\max F(f))/(\max V(f))$ for such a function.

Clearly $\max F(f)$ is attained by the use of a net consisting of one line through each row and column of points and one between every two rows and columns, together with the lines forming the boundary of R, since by our preliminary remark the net obtained by omitting any lines cannot yield a larger sum, and adding any line is extraneous as it merely introduces an additional row or column of cells each of which contributes zero to the sum. The position of the intermediate lines of the net is immaterial.

Let N, then, be such a net on R, and consider next the problem of choosing the ϵ's and $\bar{\epsilon}$'s so that $F_N(f)$ is a maximum.

Form the related matrix

$$\|a_{ij}\| = \left\| \begin{matrix} a_{1n} & a_{2n} & \cdots & a_{mn} \\ \cdot & \cdot & \cdots & \cdot \\ a_{12} & a_{22} & \cdots & a_{m2} \\ a_{11} & a_{21} & \cdots & a_{m1} \end{matrix} \right\|$$

where $a_{ij} = f(p_{ij})$. Let

$$F' = \sum_{i=1,\,j=1}^{m,\,n} \delta_i \bar{\delta}_j a_{ij} \qquad (|\delta_i| = |\bar{\delta}_j| = 1).$$

Suppose the δ's and $\bar{\delta}$'s to be so chosen that $\max F'$ is attained. To a particular

$\bar{\delta}_k$ of the sum F' there correspond two consecutive $\bar{\ }$'s of the sum $F_N(f)$; namely, those attached to the two rows of cells of N whose top and bottom edges, respectively, pass through the kth row of points p. If $\bar{\delta}_k$ is positive, let these be assigned positive and negative values respectively, while if $\bar{\delta}_k$ is negative let their signs be fixed in the reverse order. Let all the ϵ's and $\bar{\epsilon}$'s be determined in this manner.

This choice will make $F_N(f)$ assume its maximum. For it will be seen that if a particular term $\delta_i\bar{\delta}_j a_{ij}$ has the value $+1$, then the four cells of N which have the point p_{ij} in common will together contribute $+4$ to the sum $F_N(f)$, while if this term has the value -1, these cells will contribute -4; so by this choice we have $F_N(f)=4 \max F'$. Suppose now that by some other choice of ϵ's and $\bar{\epsilon}$'s we should have $F_N(f)>4 \max F'$. If for some k we have $\bar{\epsilon}_{2k+1}=\bar{\epsilon}_{2k}$, the two rows of cells of N to which these $\bar{\epsilon}$'s are attached contribute zero to the sum $F_N(f)$; and such contribution as these rows make when $\bar{\epsilon}_{2k+1}=+1$ and $\bar{\epsilon}_{2k}=-1$ is minus that which they make when the values of the $\bar{\epsilon}$'s are interchanged. Hence if by *any* choice it be possible to make $F_N(f)>4 \max F'$, there must exist *such a choice in which the ϵ's and $\bar{\epsilon}$'s occur by pairs with different signs*. But in this case we may, by reversing the process above, choose the δ's and $\bar{\delta}$'s in the sum F', and it will be seen that with this choice we have $4F'=F_N(f)>4 \max F'$, which is a contradiction. Hence $\max F_N(f)=4 \max F'$, and since, as previously remarked, we may attain the maximum of $F(f)$ by using the net N, we have

$$\max F(f) = 4 \max F'.$$

If now we denote by $\max V'$ the sum

$$\sum_{i=1,\, j=1}^{m,n} |a_{ij}|,$$

we easily see that

$$\max V(f) = 4 \max V',$$

and hence

$$\frac{\max F(f)}{\max V(f)} = \frac{\max F'}{\max V'}.$$

We proceed to show that given any $\epsilon>0$, there exists a matrix $\|a_{ij}\|$ with elements ± 1 for which $(\max F')/(\max V')$ is $<\epsilon$. This being so, we may then assert the existence of a "point-rectangle function" $f(x, y)$ for which $(\max F(f))/(\max V(f))$ is $<\epsilon$ for any preassigned $\epsilon>0$.

Consider the matrix

$$\|a_{ij}\| = \left\| \begin{matrix} a_{1n} & a_{2n} & \cdots & a_{2^{n-1},n} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ a_{11} & a_{21} & \cdots & a_{2^{n-1},1} \end{matrix} \right\| \qquad (n \text{ odd}),$$

in which all the elements of the bottom row are $+1$, and the rest of the various columns consist of the 2^{n-1} possible ordered sets of $n-1$ elements each equal to $+1$ or -1. Consider max F' for this matrix. Let the $\bar{\delta}$'s be assigned in any arbitrary way; then in order to make F' as large as possible, choose the δ's so that the total sum contributed by each column shall be positive. If this be done, we see that

1 column of elements contributes n,

n columns of elements contribute $n-2$ each,

$\dfrac{n(n-1)}{2}$ columns of elements contribute $n-4$ each,

. . . .

$\dfrac{n(n-1)(n-2)\cdots\left(\frac{n+3}{2}\right)}{((n-1)/2)!}$ columns of elements contribute 1 each,

and hence

$$F' = n + n(n-2) + \frac{n(n-1)}{2}(n-4) + \cdots + \frac{n(n-1)(n-2)\cdots\left(\frac{n+3}{2}\right)}{((n-1)/2)!}.$$

Moreover this value is independent of the choice of the $\bar{\delta}$'s, so that max F' equals this expression. Clearly we have max $V' = n2^{n-1}$. Thus for matrices of this type, we have

$$\lim_{n\to\infty} \frac{\max F'}{\max V'} = 0,$$

since the expression for $(\max F')/(\max V')$ reduces to

$$\frac{(n-1)(n-2)\cdots\left(\frac{n+1}{2}\right)}{((n-1)/2)!\,2^{n-1}},$$

which by Stirling's formula is $O(1/n^{1/2})$, and so tends to zero with $1/n$.

We now construct example

(E), a function in class F but not V. Let I, the unit square, be divided into quarter squares, and let S_1 be the upper left-hand quarter square. Next divide the lower right-hand square into quarter squares, and let S_2 be that quarter which has a common vertex with S_1, etc. We obtain in this way an infinite sequence of square subdivisions of I converging toward the point $(1, 0)$. Now in S_1 let a "point-rectangle function" be defined for which we have $\max V(f)=1$ and $\max F(f)<1/2$; thus if the set of points p_{ij} contains $n_1 2^{n_1-1}$ points, $f(p_{ij})$ is $\pm 1/(4n_1 2^{n_1-1})$ for each i and j. Similarly, in each S_j, define a "point-rectangle function" for which $\max V(f)=1$, $\max F(f)<1/2^j$. At all remaining points of I let $f(x, y)$ be zero.

It is readily seen that this function is not in V. Consider any net N on I, and let S_k be the last S_j through which lines of this net pass. By adding a sufficient number of lines to insure the largest possible contribution from each $S_j (j \leqq k)$, which cannot decrease $\max F_N(f)$, we see that for the net N we have

$$\max F_N(f) < \frac{1}{2} + \frac{1}{4} + \cdots + \frac{1}{2^k} \cdot$$

Hence $f(x, y)$ is in class F, and relation (14) is established.

Proof of (15). The first part follows from example

$$\text{(F)} \qquad f(x, y) = \begin{cases} 1 \text{ on main diagonal (through } (0, 1), (1, 0)) \text{ of I,} \\ 0 \text{ elsewhere in I,} \end{cases}$$

which is in A but not in F. The second part may be deduced from example (B).

Proof of (16). Examples (F) and (B).

Proof of (17). Examples (F) and (B).

5. **Relations concerning the extent of the common part of two or more classes.** We first establish the following relations involving a single class on the one hand and the product of two classes on the other*:

(18)	$H = A \cdot V,$	(21)	$H < A \cdot T,$
(19)	$H = V \cdot T,$	(22)	$H < P \cdot T,$
(20)	$H = P \cdot V,$	(23)	$H < A \cdot F,$

* From this list are intentionally omitted all relations such as $P > P \cdot F$, in which the class on the left appears also on the right; the inequality is definite in the light of relations (9)–(13), (15)–(17). From this and all subsequent lists all relations involving "reducible" products (such as $P \cdot A$ which reduces to A by (1), and $P \cdot V$ which reduces to H by (20)) are also omitted.

(24) $H < P \cdot F$,

(25) $H < F \cdot T$,

(26) $T > A \cdot F$,

(27) $T > P \cdot F$,

(28) $A \gtreqless P \cdot T,\ P \cdot T \ngtr A$,

(29) $V \gtreqless A \cdot T,\ A \cdot T \ngtr V$,

(30) $V \gtreqless P \cdot T,\ P \cdot T \ngtr V$,

(31) $F \gtreqless A \cdot T,\ A \cdot T \ngtr F$,

(32) $F \gtreqless P \cdot T,\ P \cdot T \ngtr F$,

(33) $V \gtreqless A \cdot F,\ A \cdot F \ngtr V$,

(34) $V \gtreqless P \cdot F,\ P \cdot F \ngtr V$,

(35) $V \gtreqless F \cdot T,\ F \cdot T \ngtr V$,

(36) $A \gtreqless F \cdot T,\ F \cdot T \ngtr A$.

Proof of (18). By (1) and (3) we have $H \leqq A \cdot V$. But a function of class V satisfies the first condition of definition H, and a function which is in A satisfies the two latter conditions of H. Hence we have $A \cdot V \leqq H$, and (18) follows.

Proof of (19). From (3) and (8) follows $H \leqq V \cdot T$. But if $f(x, y)$ is in $V \cdot T$, it satisfies the first condition of definition H, and by definition T the total variation functions $\phi(\bar{x})$ and $\psi(\bar{y})$ must be finite almost everywhere; thus we have $V \cdot T \leqq H$, and hence (19).

Proof of (20). The relation $H \leqq P \cdot V$ follows from (1) and (3). But if $f(x, y)$ is in $P \cdot V$, by Theorem 2 the functions $\phi(\bar{x})$ and $\psi(\bar{y})$ are surely finite for at least one point in their respective intervals; and as the first condition of definition H is also satisfied, we have $P \cdot V \leqq H$, and hence (20).

Proof of (21). From (1) and (8) we obtain $H \leqq A \cdot T$. From example (F), which is in $A \cdot T$ but not H, (21) is inferred.

Proof of (22). The relation $H \leqq P \cdot T$ is a consequence of (1) and (8). Example (F) then establishes (22).

Proof of (23). By (1) and (3) we have $H \leqq A \cdot F$. Then consider example (E). That function was shown to be in class F; it is, moreover, in class A. For let (x_i, y_i) be any set of points as used in definition A. Then $f(x_i, y_i)$ vanishes at all these points excepting at most those which lie within one square S_j. For this set of points we have

$$\sum \left| \Delta f(x_i, y_i) \right| \leqq 2(2^{n_j - 1} + n_j - 1)[1/(n_j 2^{n_j - 1})],$$

where $n_j 2^{n_j - 1}$ is the number of points in the array p_{ij} used to define the "point-rectangle function" in S_j. But as this expression is bounded, and indeed approaches zero with $1/n_j$, $f(x, y)$ is in A. Hence $f(x, y)$ is in $A \cdot F$, but since it is not in V it cannot be in H, from which fact (23) follows.

Proof of (24). This is implied by relation (23).

Proof of (25). By (3) and (8) we have $H \leqq F \cdot T$. Example (E) is clearly

in class T, since $\phi(\bar{x})$ and $\psi(\bar{y})$ are zero except for a denumerable set of points; and since it is also in F but not in H, we infer (25).

Proof of (26). By Theorem 3, if $f(x, y)$ is in class $A \cdot F$, $\phi(\bar{x})$ and $\psi(\bar{y})$ must be bounded and integrable in the sense of Riemann, from which we have $T \geqq A \cdot F$. Then relation (26) follows from relation (12).

Proof of (27). By Theorems 2 and 3, if $f(x, y)$ is in class $P \cdot F$, $\phi(\bar{x})$ and $\psi(\bar{y})$ are bounded and integrable in the Riemann sense, whence follows the relation $T \geqq P \cdot F$. From relation (13), relation (27) then follows.

Proof of (28). The first part follows from example (A), which is in $P \cdot T$ but not A. The second part is implied by the second of the relations (12).

Proof of (29). The first of these relations follows from example (F); the second from the first of relations (10).

Proof of (30). Example (F); relation (11).

Proof of (31). Example (F); relation (15).

Proof of (32). Example (F); relation (16).

Proof of (33). The first relation is shown by example (E), which has been proved to be in F and A, but not in V. The second part is a consequence of example (B).

Proof of (34). Example (E); relation (9).

Proof of (35). Example (E); relation (11).

Proof of (36). The second part of this relation is a consequence of (12). To establish the first part we shall now exhibit a function which is in $F \cdot T$ but not A.

As a preliminary step we define a matrix $\|a_{ij}\|$ in the following manner.* Let

$$a_{i1} = a_{1j} = 0 \qquad (i, j = 1, 2, 3, \cdots, n+1),$$

and determine the remaining elements by assigning the values of

$$\Delta_{ij} = a_{i+1,j+1} - a_{i+1,j} - a_{i,j+1} + a_{ij} \qquad (i, j = 1, 2, 3, \cdots, n)$$

to satisfy the conditions

$$|\Delta_{ij}| = 1 \qquad (i, j = 1, 2, 3, \cdots, n),$$

$$\sum_{j=1}^{n} \Delta_{ij}\Delta_{i'j} = 0 \text{ for } i' \neq i, \qquad \sum_{i=1}^{n} \Delta_{ij}\Delta_{ij'} = 0 \text{ for } j' \neq j.$$

It is known that there exist such orthogonal matrices $\|\Delta_{ij}\|$ for an infinite sequence of values of n. For such a matrix the sum

* We gratefully acknowledge our indebtedness to the late Dr. R. E. A. C. Paley for the construction of this matrix.

$$F = \sum_{i,j=1}^{n} \epsilon_i \bar{\epsilon}_j \Delta_{ij} = \sum_{i=1}^{n} \epsilon_i u_i \qquad (|\epsilon_i| = |\bar{\epsilon}_j| = 1),$$

where

$$u_i = \sum_{j=1}^{n} \Delta'_{ij}, \quad \Delta'_{ij} = \bar{\epsilon}_j \Delta_{ij},$$

is $O(n^{3/2})$, since the matrix $\|\Delta'_{ij}\|$ is also orthogonal, and by Schwarz's inequality we have $F \leqq (\sum \epsilon_i^2 \sum u_i^2)^{1/2} = n^{3/2}$.

Let $d_{ij} = a_{i,j+1} - a_{ij}$; then we have

$$d_{Ij} = \sum_{i=1}^{I} \Delta_{ij},$$

and the sum

$$\sum_{j=1}^{n} |d_{Ij}| = \sum_{j=1}^{n} \bar{\bar{\epsilon}}_j d_{Ij},$$

where

$$\bar{\bar{\epsilon}}_j = \operatorname{sgn} d_{Ij},$$

is also readily seen to be $O(n^{3/2})$ by a second application of Schwarz's inequality. This sum represents the "total variation" in the Ith row ($I=1, 2, 3, \cdots, n+1$) of $\|a_{ij}\|$. It is evident from symmetry that the total variation in the Jth column ($J=1, 2, 3, \cdots, n+1$) of $\|a_{ij}\|$ is also $O(n^{3/2})$.

Now for each j consider

$$\max_{I} |d_{Ij}| \qquad (I = 1, 2, 3, \cdots, n+1);$$

let I_j be the least value of I for which this maximum is assumed. If the numbers I_j do not increase monotonically with j, the columns of $\|\Delta_{ij}\|$ may be re-arranged so that they do; this will clearly leave undisturbed the properties of the matrix $\|\Delta_{ij}\|$ described above. Numbering the rows of both matrices $\|a_{ij}\|$ and $\|\Delta_{ij}\|$ from bottom to top we observe that the sum

$$\sum_{j=1}^{n} |d_{I_j,j}|$$

is part of what may be thought of as an Arzelà-sum for the matrix $\|a_{ij}\|$. But this sum, and therefore the maximum Arzelà-sum for the matrix, will not be $O(n^{3/2})$ if the matrix $\|\Delta_{ij}\|$ be defined thus*:

* The orthogonality of this matrix is shown by Paley, *On orthogonal matrices*, Journal of Mathematics and Physics of the Massachusetts Institute of Technology, vol. 12 (1933), pp. 311–320. That $|d_{I_j,j}|$ is not $O(n^{1/2})$ is indicated by Paley, *Note on a paper of Kolmogoroff and Menchoff*, forthcoming in the Mathematische Zeitschrift,

$$\Delta_{ij} = \begin{cases} \chi(i-j) & \text{for } i \neq j,\ i \neq 0,\ j \neq 0; \\ +1 & \text{for } i = j; \\ -1 & \text{for } i = 0 \text{ or } j = 0 \text{ but } i \neq j; \end{cases}$$

where $\chi(m)$ is a real primitive Dirichlet's character to the prime modulus p, with $p=n-1\equiv 3 \bmod 4$.

We may now construct example

(G), a function in $F\cdot T$ but not A. Let $\{S_k\}$ $(k=1, 2, 3, \cdots)$ be an infinite sequence of square subdivisions of the unit square I similar to that employed in example (E) but converging toward the point (1, 1). By the above discussion there exists a matrix $\|a_{ij}^{(k)}\|$, of n_k rows and n_k columns, for which F and the total variation in each row and column is $<1/2^k$ while the maximum Arzelà-sum is >1. In S_k $(k=1, 2, 3, \cdots)$ let $p_{ij}^{(k)}$ be a square array of n_k^2 points, with rows and columns equally spaced. The points $p_{ij}^{(k)}$ then determine a set of square cells. In the cell whose vertices are $p_{ij}^{(k)}$, $p_{i,j+1}^{(k)}$, $p_{i+1,j}^{(k)}$ and $p_{i+1,j+1}^{(k)}$, including its boundary, let $f(x, y)=a_{ij}^{(k)}$ at each point except along the top and right-hand sides. At all other points of I let $f(x, y)=0$. The function $f(x, y)$ is then in both F and T but is not in A.

We list the following additional relations and indicate briefly the proof of each*:

(37) $A\cdot T < P\cdot T$, (38) $A\cdot T \gtreqless\!\!\!\!/\ F\cdot T, F\cdot T \not> A\cdot T$, (39) $A\cdot F < A\cdot T$,

(40) $A\cdot F < P\cdot T$, (41) $A\cdot F<F\cdot T$, (42) $P\cdot F < P\cdot T$,

(43) $A\cdot F = A\cdot F\cdot T$, (44) $P\cdot F = P\cdot F\cdot T$.

Proof of (37). Relations (1) and (28).
Proof of (38). Relations (36) and (31).
Proof of (39). Relations (26) and (31).
Proof of (40). Relations (1), (26), and (32).
Proof of (41). Relations (26) and (36).
Proof of (42). Relations (27) and (32).
Proof of (43). Relation (26).
Proof of (44). Relation (27).

6. **Relations between classes when only bounded functions are admitted to consideration.** Reasoning similar to that of the preceding sections readily shows that each of the forty-four relations given above remains valid if bounded functions alone are considered. We thus have the further results†:

---

* Relations (43) and (44), together with (19), show that there are no irreducible products of three classes; hence such products need no further consideration.

† Numbers are used here and later to correspond with those of similar relations given above.

(3b) $V \cdot B > H$,

(8b) $T \cdot B > H$,

(9b) $V \cdot B \not\gtrless P,\ P \not> V \cdot B$,

(10b) $A \not\gtrless V \cdot B,\ V \cdot B \not> A$,

(11b) $V \cdot B \not\gtrless T \cdot B,\ T \cdot B \not> V \cdot B$,

(12b) $A \not\gtrless T \cdot B,\ T \cdot B \not> A$,

and so on.

7. **Relations between classes when only continuous functions are admitted.** We establish the following set[*] of relations:

(1c) $T \cdot C > P \cdot C > A \cdot C > H \cdot C$,

(9c) $P \cdot C \not\gtrless V \cdot C,\ V \cdot C \not> P \cdot C$,

(10c) $A \cdot C \not\gtrless V \cdot C,\ V \cdot C \not> A \cdot C$,

(11c) $T \cdot C \not\gtrless V \cdot C,\ V \cdot C \not> T \cdot C$,

(14c) $F \cdot C > V \cdot C$,

(15c) $A \cdot C \not\gtrless F \cdot C, F \cdot C \not> A \cdot C$,

(16c) $P \cdot C \not\gtrless F \cdot C, F \cdot C \not> P \cdot C$,

(17c) $T \cdot C \not\gtrless F \cdot C, F \cdot C \not> T \cdot C$,

(23c) $H \cdot C < A \cdot F \cdot C$,

(24c) $H \cdot C < P \cdot F \cdot C$,

(25c) $H \cdot C < F \cdot T \cdot C$,

(33c) $V \cdot C \not\gtrless A \cdot F \cdot C, A \cdot F \cdot C \not> V \cdot C$,

(34c) $V \cdot C \not\gtrless P \cdot F \cdot C,\ P \cdot F \cdot C \not> V \cdot C$,

(35c) $V \cdot C \not\gtrless F \cdot T \cdot C, F \cdot T \cdot C \not> V \cdot C$.

The proof of (1c) will be given in three parts.

Proof of $A \cdot C > H \cdot C$. This relation was established by Küstermann, loc. cit., who gave an example of a continuous function which is in class A but not H; a simpler example is given by Hahn, loc. cit. The following function, example

(H), will be found to exhibit the same property; moreover one may easily determine whether it is in classes F and T. Let $S_1, S_2, S_3, \cdots$ be an infinite sequence of square subdivisions of I converging toward the point (1, 0) as defined in example (E). In each S_j let $f(x, y)$ be defined by the surface of a regular square pyramid whose base is S_j and height $1/j$, and let $f(x, y)$ vanish over the rest of I. Then $f(x, y)$ is continuous; it is not in V and hence not in H. For if a net N be defined whose lines consist of the lines through the sides of the squares $S_j (j=1, 2, 3, \cdots, k)$ and lines horizontally and vertically through the centers of these squares, for this net we have

$$V_N(f) = 4 \sum_{n=1}^{k} 1/n,$$

which may be arbitrarily large. But $f(x, y)$ is in class A. For if (x_i, y_i) be any set of points as used in the Arzelà definition, $f(x_i, y_i)$ vanishes except at such points as lie within one square S_j, and we have $\sum |\Delta f(x_i, y_i)| \leqq 2$.

[*] The correspondents of certain earlier relations do not appear here, since $T \cdot C$ includes both $P \cdot C$ and $A \cdot C$, whereas T includes neither P nor A. We omit also all relations such as $T \cdot C > T \cdot F \cdot C$, in which the class on the left (other than C) appears also on the right; in such relations the inequality is always definite by virtue of the relations (9c)–(11c) and (15c)–(17c).

This function is in class T, since $\phi(\bar{x})$ and $\psi(\bar{y})$ are continuous. It is not in class F; for if a net N as defined in the preceding paragraph be used, the ϵ_i's and $\bar{\epsilon}_j$'s can be chosen so that $F_N(f)=V_N(f)$, which is arbitrarily large.

Proof of $P\cdot C>A\cdot C$. We clearly have $P\cdot C\geqq A\cdot C$. To remove the possibility of equality consider example

(I), a function defined in I in precisely the same manner as example (H) except that the sequence of subsquares $\{S_i\}$ shall in this case converge toward the point (1, 1); i.e., example (I) is obtained from example (H) by changing the position of the x- and y-axes. As the function was in class P before the change, the new function is clearly in that class also; it is easily seen to be in class T but not in classes H, A, V, or F.

Proof of $T\cdot C>P\cdot C$. We first establish the relation $T\cdot C\geqq P\cdot C$.

Assume $f(x, y)$ to be of class $P\cdot C$ in R and suppose $\sum_{\nu=1}^{n^2}\omega_\nu'/n<M$. Let the sequence of functions $\{\phi_n(\bar{x})\}$ be defined in the interval $a\leqq\bar{x}\leqq b$ as follows: for a fixed n and $\bar{x}$,

$$\phi_n(\bar{x}) = \sum_{i=1}^{2^n}\left| f(\bar{x}, y_i) - f(\bar{x}, y_{i-1})\right|,$$

where $y_0=c$, $y_n=d$, and $y_i-y_{i-1}=(d-c)/2^n$ $(i=1, 2, 3, \cdots, 2^n)$. Then each $\phi_n(\bar{x})$ is continuous, and we have

$$\text{(d)} \qquad \lim_{n\to\infty}\phi_n(\bar{x}) = \phi(\bar{x});$$

moreover the sequence $\{\phi_n(\bar{x})\}$ is positive and non-decreasing. Hence* we have

$$\lim_{n\to\infty}\int\phi_n(x)\,dx = \int\phi(x)dx.$$

Now for any fixed n let the symmetrical net N of 2^{2n} congruent cells be considered, and let I_j $(j=1, 2, 3, \cdots, 2^n)$ be the jth of the 2^n equal parts into which N divides the interval $a\leqq x\leqq b$; then we have

$$\int\phi_n(x)dx = \sum_{j=1}^{2^n}\int_{I_j}\phi_n(x)dx.$$

Let B_j denote the least upper bound of $\phi_n(\bar{x})$ in I_j; then

$$\begin{aligned}\int\phi_n(x)dx &\leqq \sum_{j=1}^{2^n}\int_{I_j}B_j dx\\ &= [(b-a)/2^n]\sum_{j=1}^{2^n}B_j.\end{aligned}$$

* See, for example, de la Vallée Poussin, *Cours d'Analyse Infinitésimale*, vol. 1, Paris, 1914, p. 264, Theorem III.

But B_j is at most equal to the sum of the oscillations ω_ν' in the jth column of cells of N, whence

$$\begin{aligned}[(b-a)/2^n]\sum_{j=1}^{2^n} B_j &\leqq (b-a)\sum_{\nu=1}^{2^{2n}}\omega_\nu'/2^n \\ &< M(b-a).\end{aligned}$$

Thus for all n, $\int\phi_n(x)dx$ is $<M(b-a)$ and consequently, by (d), $\phi(\bar{x})$ is summable. Since the same reasoning holds for $\psi(\bar{y})$, the relation $T\cdot C\geqq P\cdot C$ is proved.

We now construct example

(J), a function $f(x, y)$ in class $T\cdot C$ but not P, thus establishing the relation $T\cdot C>P\cdot C$. To this end we employ a result of Tonelli,* that if $f(x, y)$ is continuous and if the surface $z=f(x, y)$ is of finite area†, $f(x, y)$ is in class T.

Let N_j $(j=2, 3, 4, \cdots)$ be the net which divides I, the unit square, into $2^{2(j-1)}$ equal subsquares Q_{ji} $(i=1, 2, 3, \cdots, 2^{2(j-1)})$. Thus N_{j+1} divides each subsquare Q_{ji} of N_j into four equal subsquares. We shall define the function $f(x, y)$ over I by a surface Z which will in turn be defined as the limit of a sequence $\{Z_j\}$ of polyhedral surfaces over I, Z_j corresponding to the net N_j.

Let Z_1 be a regular pyramid Δ_1 whose base is I and altitude 1. Its surface area may be denoted by $S/2$.

Let Z_2 be identical with Z_1 except over the squares of a set Q_2' concentric with the squares Q_2 of N_2. Let a second set of smaller concentric squares Q_2'' be chosen. The squares of Q_2' may be taken as small as desired, and, these having been chosen, the squares of Q_2'' may be selected as small as desired. Limitations on their size are presently to be imposed.

As a first limitation on Q_2' let the oscillation of Z_1 be less than $\frac{1}{2}$ in each square of Q_2'. Within Q_{2i}'' (where this is the square of Q_2'' interior to Q_{2i}', in turn interior to Q_{2i}) define Z_2 as a regular pyramid Δ_{2i} of altitude $\frac{1}{2}$ and with base in a horizontal plane. The plane of the base of Δ_{2i} may be so chosen that Δ_{2i} lies wholly between the two horizontal planes through the lowest and highest points of Z_1.

Figure 1 is intended to indicate a top elevation of the part of the surface Z_2 now being described. $ABCD$ is the space quadrilateral on Z_1 whose projection on the xy-plane is Q_{2i}; $A'B'C'D'$ is the space quadrilateral on Z_1 whose projection is Q_{2i}'; and $A''B''C''D''$ is the base of the pyramid Δ_{2i} whose projection is Q_{2i}''. Let a, b, c, and d be the mid-points of the sides of $A''B''C''D''$. Then plane triangles may be interpolated between the space quadrilateral

* See Tonelli, loc. cit.

† In the sense of Lebesgue.

$A'B'C'D'$ and the base of Δ_{2i}; these triangles are $A'A''a$, $A'aB'$, $aB''B'$, etc. The plane triangles thus interpolated we use to define the part of Z_2 standing over the region between the two squares Q'_{2i} and Q''_{2i}; Z_2 so defined is continuous within Q'_{2i}, and hence throughout I. The position of the plane of the base of Δ_{2i} is further restricted merely by the condition that the oscillation of Z_2 in Q'_{2i} (which by the presence of the pyramid Δ_{2i} is not less than $\frac{1}{2}$) shall be $\frac{1}{2}$. Evidently, by decreasing the size of the squares Q'_{2i} and Q''_{2i}, we may make the

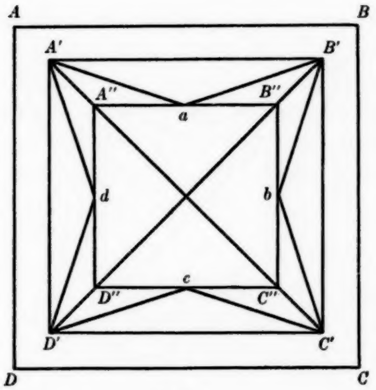


Fig. 1

surface area of Z_2 within Q'_{2i} as small as we wish; hence we may impose the final limitation upon the size of the squares, that the resulting area of Z_2 shall not exceed $S(\frac{1}{2}+\frac{1}{4})$. To provide for further subdivision we require that the lengths of the sides of the squares in Q_2, Q_2', and Q_2'' be relatively incommensurable.

Each succeeding surface Z_p is defined by means of the surface Z_{p-1} in a similar manner. Let Z_p be identical with Z_{p-1} except over the squares of a set Q_p' concentric with the squares Q_p of N_p. Let Q'_{pi} be chosen sufficiently small so that its perimeter does not intersect the perimeter of any previously chosen Q'_{ij} or Q''_{ij}. Let a second set of smaller concentric squares Q_p'' be chosen.

As the next limitation on Q_p' let the oscillation of Z_{p-1} be less than $1/p$ in each square of Q_p'. Within Q''_{pi} (where this is the square of Q_p'' interior to Q'_{pi}, in turn interior to Q_{pi}) define Z_p as a regular pyramid Δ_{pi} of altitude $1/p$ and with base in a horizontal plane. Q'_{pi} lies entirely within some smallest previously chosen Q'_{ij} (which may be I itself), Q'_{mn}. The plane of the base of Δ_{pi} may be so chosen that Δ_{pi} lies wholly between the two horizontal planes through the highest and lowest points of Z_{p-1} in Q'_{mn}.

Figure 2 is intended to indicate a top elevation of the part of the surface Z_p now being described. $ABCD$ is the space polygon on Z_{p-1} whose projection on the xy-plane is Q_{pi}; $A'p_1p_2 \cdots B' \cdots C' \cdots D' \cdots$ is the space polygon on Z_{p-1} whose projection is Q'_{pi}; and $A''B''C''D''$ is the base of the pyramid Δ_{pi} whose projection is Q''_{pi}. Let a, b, c, and d be the mid-points of the sides of $A''B''C''D''$. Then plane triangles may be interpolated between the space polygon $A'p_1p_2 \cdots B' \cdots C' \cdots D' \cdots$ and the base of Δ_{pi}; these triangles

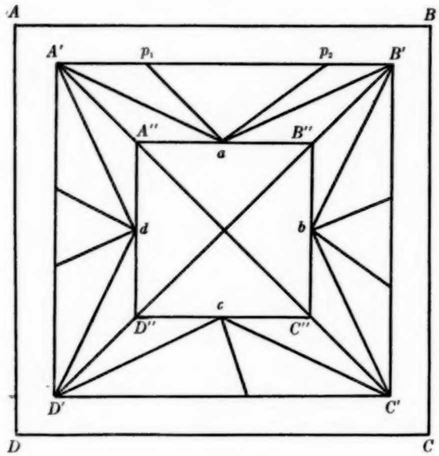


Fig. 2

are $A'A''a$, $A'ap_1$, p_1ap_2, etc. The plane triangles thus interpolated we use to define the part of Z_p standing over the region between the two squares Q'_p and Q''_{pi}; Z_p is then continuous within Q'_{pi}, and hence throughout I. The position of the plane of the base of Δ_{pi} is further restricted merely by the condition that the oscillation of Z_p in Q'_{pi} (which by the presence of the pyramid Δ_{pi} is at least $1/p$) shall be $1/p$. The final limitation upon the size of the squares Q'_{pi} and Q''_{pi} is that the resulting surface area of Z_p shall not exceed $S(\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+ \cdots +1/2^p)$. In order to show that this result may be effected, we need only prove that the total surface area of Z_p thus defined within Q'_{pi} may be made arbitrarily small by choosing the squares Q'_{pi} and Q''_{pi} sufficiently small.

Let any $\epsilon>0$ be given. Then clearly there exists a δ_1 such that if the side of Q''_{pi} be taken less than δ_1, the surface area S' of the pyramid Δ_{pi} will be less than $\epsilon/2$. Now consider S'', the total surface area of the plane triangles between Q'_{pi} and Q''_{pi}. Of these triangles eight have the property that each has

one side which coincides with half of a side of the base of Δ_{pi}; moreover the length of each of its other sides is bounded by $((d/2)^2+(1/p)^2)^{1/2}$, where d is the length of the diagonal of Q'_{pi}; whence we may assert that there exists a δ_2 such that if the side of Q''_{pi} be taken less than δ_2 in length, the surface area of these eight triangles will be less than $\epsilon/4$. There remain to be considered the rest of the triangles which contribute to S''. Each of these has one side whose length is bounded by $(\omega^2+l^2)^{1/2}$, where ω is the oscillation of Z_{p-1} in Q'_{pi} and l is the length of a side of Q'_{pi}; likewise each of its other sides is bounded by $((d/2)^2+(1/p)^2)^{1/2}$. Moreover the number of these triangles is limited, since the surface Z_{p-1} consists of a finite number of plane pieces. Now by taking Q'_{pi} sufficiently small we may make ω and l, and consequently one side of each of these triangles, as small as we please; hence there exists a δ_3 such that if the side of Q'_{pi} be taken less than δ_3, the combined areas of these remaining triangles will be less than $\epsilon/4$. If, then, we require that the side of Q'_{pi} be less than δ_3, and the side of Q''_{pi} be less than δ_1 and δ_2, the total area of the part of Z_p within Q'_{pi} will be less than ϵ.

Finally, to provide for further subdivision, we take the lengths of the sides of the squares in Q_p, Q_p', and Q_p'' relatively incommensurable.

Then if P is any point of I, and h_j denotes the height of Z_j over P, the sequence $\{h_j\}$ approaches a limit as j increases indefinitely. For if P does not lie within an infinite number of squares Q'_{ij}, all the h_j's are equal for sufficiently large j. If P does lie within an infinite sequence of such squares, we have $|h_j-h_p|<1/p$ for all $j>p$, and so the sequence $\{h_j\}$ converges. Let the surface Z be defined as the limit of the sequence $\{Z_j\}$.

Inasmuch as each Z_j is continuous, and the sequence $\{Z_j\}$ converges uniformly, the surface Z is continuous and defines a continuous function $f(x, y)$ over I. Moreover, as Z may be approximated arbitrarily closely by one of the sequence $\{Z_j\}$ of polyhedral surfaces, each of which is of area less than S, the area of Z does not exceed S; hence $f(x, y)$ is in class T. But for each net N_j we have

$$\sum_{\nu=1}^{n^2}\omega_\nu'/n \geqq [2^{2(j-1)}(1/j)]/2^{j-1} = 2^{j-1}/j,$$

and as the latter quantity increases indefinitely with j, the function $f(x, y)$ is not in class P.

Proof of (9c). Example (B); example (I).

Proof of (10c). Example (B); example (H).

Proof of (11c). Example (B); example (H).

Proof of (14c). Example (E) has already been given to exhibit a function which is in class F but not in class V. We now show how this example

may be modified so as to be continuous without otherwise essentially altering its character.

Consider a "point-rectangle function" $f(x, y)$ such as is used in example (E), with $|f|=1$ on the array of $n2^{n-1}$ points p_{ij} in R and with $(\max F(f))/(\max V(f))<\epsilon$. Surround each point p_{ij} by a square Q_{ij} with sides parallel to the axes and with p_{ij} as center; all these squares are taken equal in size and small enough so that they do not abut or overlap.

Let $f'(x, y)$ be defined on R as follows: within each Q_{ij} let $f'(x, y)$ be defined by the surface of a regular pyramid whose base is Q_{ij} and whose height is $f(p_{ij})$; let $f'(x, y)=0$ at all other points of R. This function is continuous on R, and for it also we have

$$(\max F(f'))/(\max V(f')) < \epsilon. \tag{e}$$

For the following inequalities are easily seen to hold:

$$\max F(f') \geqq \max F(f), \quad \max V(f') \geqq \max V(f);$$

and we shall show that max $F(f')$ does not exceed max $F(f)$, whence (e) will follow.

Let N be any net on R. Construct a second net N' by adding lines to N as follows. Add the horizontal lines through the center and upper and lower sides of Q_{11}, and the corresponding vertical lines. If a horizontal line l of N passes through Q_{11}, add to N the horizontal line l' so that p_{11} is equidistant from l and l', and also the vertical lines l'' and l''' at the same distance from p_{11}. For each Q_{ij} add to N four lines bearing the same relation to it that l, l', l'' and l''' bear to Q_{11}. Let this construction be carried out for each horizontal and vertical line through Q_{11}, and then the process repeated for every other Q_{ij}.

The net N' thus defined is symmetric; each Q_{ij} is divided by N' in precisely the same way into t^2 rectangular subregions which are in general not square except for those along the diagonals, and for each of which, excepting those along the diagonals, the difference $\Delta_{11}f'(x_i,y_j)$ vanishes. For each of the squares along the diagonals the difference is in absolute value equal to one of the $t/2$ values $a, b, c, \cdots, k$, where $a+b+c+\cdots+k=1$. The values of t and of the numbers $a, b, c, \cdots, k$ depend upon the net N'. The situation for a particular Q_{ij} may be represented as in Figure 3, where the number within each cell is the value of the difference $\Delta_{11}f'$ for that cell. The figure represents a Q_{ij} when $f'(p_{ij})=+1$ and $t/2=4$. It will be seen that the ϵ_i's and $\bar{\epsilon}_j$'s in the sum $F_{N'}(f')$ which attach to the cells yielding $\pm a$ may be chosen independently of those which attach to the cells yielding $\pm b$, etc., and that

the maximum contribution obtainable from the cells yielding $\pm a$ is $a\cdot\max F(f)$; from those yielding $\pm b$, $b\cdot\max F(f)$; etc. Hence we have

$$\max F_{N'}(f') = a\cdot\max F(f) + b\cdot\max F(f) + \cdots + k\cdot\max F(f) = \max F(f).$$

But as $\max F_N(f')$ is $\leqq \max F_{N'}(f')$ (since N' was obtained from N by adding lines) and N was an arbitrary net, we conclude that $\max F(f') = \max F(f)$, which was to be proved.

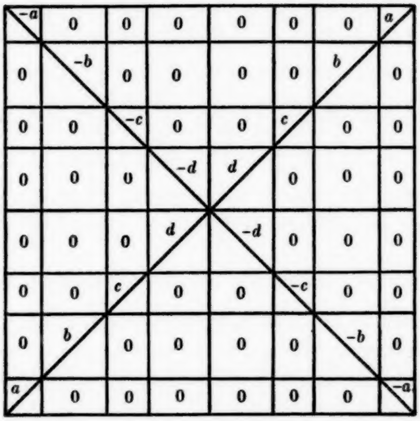


Fig. 3

(K) Since a "point-rectangle function" may be made continuous while retaining the same values for $\max F(f)$ and $\max V(f)$, we may clearly construct example (K) by modifying example (E) in this way, and so obtain a continuous function which is in class F but not V.

Proof of (15c). Example (B); example (H).

Proof of (16c). Example (B); relation (15c).

Proof of (17c). Example (B); relations (5) and (15c).

Proof of (23c). Example (K) is readily seen to be in class $A\cdot F\cdot C$, but not in V and hence not in H. Since we have $H\cdot C \leqq A\cdot F\cdot C$, (23c) follows.

Proof of (24c). Relation (23c).

Proof of (25c). Relation (23c).

Proof of (33c). Example (K); example (B).

Proof of (34c). Relation (33c); example (B).

Proof of (35c). Relation (33c); example (B).

8. **Open questions.** The following is a complete list of pairs of classes the

relations between which are not yet fully determined; in each case we give in parentheses a partial determination of the relation, with a reason therefor.

(45) $P, F\cdot T$ $(P \lesseqqgtr F\cdot T$ by (13)),

(46) $A, P\cdot F$ $(A \lesseqqgtr P\cdot F$ by (15)),

(47) $A\cdot F, P\cdot F$ $(A\cdot F \leqq P\cdot F$ by (1)),

(48) $A\cdot T, P\cdot F$ $(A\cdot T \lesseqqgtr P\cdot F$ by (31)),

(49) $P\cdot F, F\cdot T$ $(P\cdot F \leqq F\cdot T$ by (27)),

(50) $P\cdot T, F\cdot T$ $(P\cdot T \lesseqqgtr F\cdot T$ by (32)),

(36c) $A\cdot C, F\cdot T\cdot C$ $(A\cdot C \lesseqqgtr F\cdot T\cdot C$ by (15c)),

(41c) $A\cdot F\cdot C, F\cdot T\cdot C$ $(A\cdot F\cdot C \leqq F\cdot T\cdot C$ by (1c)),

(45c) $P\cdot C, F\cdot T\cdot C$ $(P\cdot C \lesseqqgtr F\cdot T\cdot C$ by (16c)),

(46c) $A\cdot C, P\cdot F\cdot C$ $(A\cdot C \lesseqqgtr P\cdot F\cdot C$ by (15c)),

(47c) $A\cdot F\cdot C, P\cdot F\cdot C$ $(A\cdot F\cdot C \leqq P\cdot F\cdot C$ by (1)),

(49c) $P\cdot F\cdot C, F\cdot T\cdot C$ $(P\cdot F\cdot C \leqq F\cdot T\cdot C$ by (1c)).

The relations still to be determined present some interesting, but probably not simple, problems. We would hazard no conjecture concerning their nature except in the case of (36c) and (41c), the first of which is probably an overlapping relation and the second a definite inequality; that such is the case could be established by modifying example (G) so as to be continuous while preserving its other properties. We have no doubt that this modification is possible, but see no way to do it easily.

Brown University,
Providence, R.I.

THE GENERAL WEB OF ALGEBRAIC SURFACES OF ORDER n AND THE INVOLUTION DEFINED BY IT*

BY

TEMPLE R. HOLLCROFT

1. **Introduction.** Webs of quadric surfaces have been studied quite extensively†, but, except for finding the characteristics of its jacobian‡ the general web has not been treated for $n>2$.

The surfaces of a web are in (1, 1) correspondence with the planes of three-space. In the case of a general web of surfaces of order n, this correspondence establishes a space involution of order n^3. For $n=2$, this involution has been treated by Snyder and Sharpe.§

In the present paper, the properties of both the general web of algebraic surfaces of order n and the space involution associated with such a web are obtained.

2. **The web.** The equation of a web of algebraic surfaces is

$$\sum \lambda_i f_i = 0 \qquad (i = 1, 2, 3, 4),$$

in which the λ_i are homogeneous parameters and the f_i are homogeneous, algebraic functions of order n in the variables x_1, x_2, x_3, x_4. The general web treated in this paper is such that the coefficients in each of the four f_i defining the web are unrestricted, that is, the f_i represent non-singular surfaces and the web has no basis points or curves.

The jacobian J of a web of surfaces‖ is the locus of nodes¶ and also of contacts of surfaces of the web. It is a surface of order $4(n-1)$. For a general web, J has no singularities and is, therefore, of genus

$$D = \tfrac{1}{3}(2n-3)(4n-5)(4n-7).$$

The characteristics of J result immediately from the Cayley formulas for the characteristics of a non-singular surface.**

* Presented to the Society, December 29, 1932; received by the editors February 7, 1933.

† Pascal, Repertorium der höheren Mathematik, vol. II_3 (1922), pp. 629–631. Encyklopädie der Mathematischen Wissenschaften, vol. III_2, pp. 250–254.

‡ Pascal, loc. cit., p. 680.

§ Virgil Snyder and F. R. Sharpe, *Space involutions defined by a web of quadrics*, these Transactions, vol. 19 (1918), pp. 275–290.

‖ As only algebraic curves and surfaces are treated in this paper, the adjective "algebraic" will be omitted from this point on.

¶ The term "node" is used to mean an entirely general conic node.

** Pascal, loc. cit., pp. 696–697.

Since a web involves three essential parameters, the surfaces of the web may have contacts or singularities associated with one, two or three invariants. The following singularities on one surface and contacts of surfaces are associated with one, two or three invariants:

I. One invariant. (a) One node or one contact.

II. Two invariants. (a) Two nodes or two contacts. (b) One binode or one stationary contact.

III. Three invariants. (a) Three nodes or three contacts. (b) One binode and one node or one stationary and one simple contact. (c) One special binode B_4 whose axis has four-point contact with the surface or one contact such that the curve of intersection has a tacnode at the point of contact.

One invariant defines a doubly infinite system of surfaces of the web, the locus of whose singularities or contacts associated with the given invariant is the jacobian surface J. Thus J is the locus of nodes and contacts belonging to both doubly infinite systems of surfaces of the web.

Two invariants define a singly infinite system of surfaces of the web, the loci of whose singularities or contacts associated with the two given invariants are curves on J. The characteristics of these four curves are obtained in §§6 and 7.

Three invariants define a finite system of surfaces of the web whose singularities or contacts associated with the three given invariants lie at certain intersections or contacts of the above curves on J. The positions and numbers of these for each of the six finite systems are also found in §§6 and 7.

3. **The involution defined by the web.*** The (1, 1) correspondence existing between the surfaces of the web and the planes of the three-space (y) is defined by the equations

$$\rho y_i = f_i \qquad (i = 1, 2, 3, 4).$$

To a line of (y) corresponds a space curve of order n^2 and genus $(n-1)\cdot(n^2-n-1)$, the basis curve of a pencil of surfaces of the web. The surfaces of this pencil are the images of the planes of (y) belonging to a pencil whose axis is the given line.

Conversely, the complete image of a surface of the web or of the basis curve of a pencil of surfaces is the corresponding plane or line of (y) respectively counted n^3 times.

To a bundle of planes through a point P of (y) corresponds a net of surfaces of the web with n^3 basis points. These n^3 basis points are all images of P. Conversely, to each of these n^3 points of (x) corresponds the given point P

* The results of this section are chiefly generalizations of those obtained by Snyder and Sharpe (loc. cit.) for $n=2$.

of (y). The unique correspondence of planes of (y) to surfaces of the web therefore establishes an involution of order n^3 between the spaces (x) and (y).

The locus of the points of (y), two of whose n^3 image points coincide, is a surface L called the branch-point surface of the transformation. The corresponding locus of coincidences is J. L and J are in $(1, 1)$ correspondence. The order of L is the order of the tact-invariant of two surfaces of the web, which is $4n^2(n-1)$.

The complete image of L is J counted twice and a residual surface R of order $4(n-1)(n^3-2)$. To a point of L correspond two coincident points of J and n^3-2 distinct points of R.

The complete image of J is L. The complete image of R is L counted n^3-2 times.

The image of a plane π of (x) is a rational surface s of order n^2 in (y), whose only singularity is a nodal curve. The image of s is the plane π and a residual surface s_1 of order n^3-1. The plane π meets J in a curve of order $4(n-1)$ through which s_1 passes. The residual intersection of π and s_1 is a plane curve of order $(n-1)(n^2+n-3)$, the image of the nodal curve of s. The nodal curve of s is, therefore, of order $\frac{1}{2}n(n-1)(n^2+n-3)$ and since s is rational, the rank of its nodal curve is $(n-1)(n^2-3)(2n^3+2n^2-7n+2)/3$.

The surface s intersects L in a curve of order $4n^4(n-1)$, consisting of a contact curve of order $4n(n-1)$, the image of the intersection curve of π and J; and an intersection curve of order $4n(n-1)(n^3-2)$, the image of the plane section of R by π.

The images of linear systems of planes of (x) do not form linear systems of surfaces in (y). The image of a pencil of planes of (x) is a singly infinite, non-linear system of surfaces in (y) of order n^2. This system of surfaces has in common a curve of order n, the image of the line which is the axis of the pencil of planes of (x). This curve C_n is rational and has $\frac{1}{2}(n-1)(n-2)$ apparent double points. Any two image surfaces of planes of the pencil intersect in C_n and a residual curve C' of order $n(n^3-1)$ and genus $n(3n^3-5n^2-4n+7)$ with $\frac{1}{2}(n-1)[n(n-1)(n+1)^2(n^3-2)-n-2]$ apparent double points. C' also has $n(n-1)(n^2-1)(n^2+n-3)$ nodes at its intersections with the two nodal curves of the two surfaces.

4. **Nets contained in the web.** The points of (y), each considered as bearing a bundle of planes, determine a triple infinity of nets of surfaces of the web.

Consider the net F of surfaces corresponding to the bundle of planes through an arbitrary point P of (y). The properties of F are uniquely associated with certain characteristics of the branch-point curve L_1 of a plane π_1 whose lines are in $(1, 1)$ correspondence with the surfaces of F. The follow-

ing characteristics of L_1 (which will be used in the next section to obtain the characteristics of the branch-point surface L of the web) result on setting $i=3$ in the formulas for a net of hypersurfaces in i dimensions*:

$$n_1 = 6n(n-1)^2;$$

$$m_1 = 4(n-1)^3;$$

$$\delta_1 = (n-1)^2[18n^2(n-1)^2 - 59n + 74];$$

$$\kappa_1 = 12(n-1)^2(3n-4);$$

$$\tau_1 = 2(n-1)^2(n-2)(4n^3 - 8n^2 + 8n - 25);$$

$$\iota_1 = 30(n-1)^2(n-2).$$

The jacobian curve J_1 of the net is of order $6(n-1)^2$ and is the locus of both nodes and contacts of surfaces of the net. J is also the coincidence curve of the transformation. To nodes δ_1, cusps κ_1, bitangents τ_1, stationary tangents ι_1 of L_1 correspond uniquely and respectively surfaces of F with two contacts, one stationary contact, two nodes, one binode.

The surfaces of F that have a node are surfaces of the web and therefore all these nodes lie on the surface J. The respective jacobian curves of the ∞^3 nets contained in the web form a triply infinite linear system of curves on J, that is, the jacobians of the nets of the web build a web of curves on the jacobian of the web.

Two points P and P' of (y) determine the line PP' carrying an axial pencil whose planes are common to the two bundles on P and P'. Corresponding to the two bundles and to their common axial pencil, there are in (x) two sets of surfaces with a pencil of surfaces in common. A pencil of surfaces contains $4(n-1)^3$ surfaces with a node. These surfaces belong to both nets and therefore their $4(n-1)^3$ nodes lie on both jacobian curves. Therefore, in the web of jacobian curves on J, any two jacobian curves intersect in $4(n-1)^3$ points.

To a bundle of lines on any point P of (y) corresponds a net of curves of order n^2 and genus $(n-1)(n^2-n-1)$ with n^3 basis points. These curves are the intersections of the surfaces of the net corresponding to the bundle of planes on P. The nodes of this net of curves lie at contacts of surfaces of the associated net. Any such net of curves has, therefore, the same jacobian curve J_1 as the associated net of surfaces.

5. **The characteristics of L.** The characteristics of the branch-point surface L will be represented by the following symbols:

$N[n']$ order [class];

$a[a']$ order of tangent cone [class of plane section];

* Hollcroft, *Nets of manifolds in i dimensions*, Annali di Matematica, (4), vol. 5 (1927–28), p. 265.

$\kappa'[\delta']$ number of inflections [bitangents] of plane section;
$\kappa[\delta]$ number of cuspidal [nodal] lines of tangent cone;
$b[c]$ order of nodal [cuspidal] curve;
$b'[c']$ class of bitangential [spinodal] developable;
$q[q']$ class [order] of nodal curve [bitangential developable];
$r[r']$ class [order] of cuspidal curve [spinodal developable];
$\beta[\gamma](i)$ number of intersections of nodal and cuspidal curves which are cusps on cuspidal [nodal] (neither) curve;
$\beta'[\gamma'](i')$ number of common planes of bitangential and spinodal developables which are stationary on the spinodal [bitangential] (neither) developable;
$t[t']$ number of triple points [planes] of nodal curve [bitangential developable];
$\rho'[\sigma'](\phi')$ order of bitangential [spinodal] (flecnodal) curve;
$\rho[\sigma]$ class of nodal [cuspidal] developable.

The order of L was found in §3. The class of L is the order of the discriminant of a pencil of surfaces of the web, which is $4(n-1)^3$.

The characteristics of the tangent cone to L from any point will now be obtained.

To a bundle of planes of (y) on an arbitrary point P corresponds a net F of surfaces of the web. Of the ∞^2 planes through P, a single infinity are tangent to L, enveloping the tangent cone to L from P. The planes through P tangent to L correspond uniquely to the single infinity of surfaces of the net that have a node. The planes enveloping the tangent cone to L from P are, therefore, in (1, 1) correspondence with the points of the jacobian curve J_1 of F.

The plane section L_1 of this tangent cone made by any plane π_1 not through P is enveloped by lines which are sections by π_1 of the enveloping tangent planes of the cone. The lines enveloping L_1 are thus in (1, 1) correspondence with the points of J_1. Therefore, since the lines of π_1 (sections by π_1 of the planes of the bundle) are in (1, 1) correspondence with the surfaces of F, the plane section L_1 of the tangent cone is the branch-point curve in the transformation of the lines of π_1 into the surfaces of F. From the preceding section, the order of L_1 is $6n(n-1)^2$ which is, therefore, the order a of the tangent cone to L from any point.

The number of bitangent [stationary] planes from P to L is the class b' $[c']$ of the bitangential [spinodal] developable of L. These planes are cut by π_1 in bitangents [stationary tangents] to the plane curve L_1. The number of bitangents τ_1 and stationary tangents ι_1 of L_1 were obtained in the preceding section. These values of τ_1 and ι_1 are, therefore, the values of b' and c' respectively.

The number of nodes [cusps] of the plane section L_1 is the number of nodal [cuspidal] generators $\delta[\kappa]$ of the tangent cone. The number of nodes δ_1 and cusps κ_1 of L_1 were obtained in the preceding section. These are the values of δ and κ respectively.

Since, for any algebraic surface, $a=a'$, the class of a plane section of L is $6n(n-1)^2$.

To find the genus of a plane section of L, consider the plane section L' of L made by any plane π'. L', of order $4n^2(n-1)$, has for its image in (x) the space curve J' which is the complete intersection of J and the surface of the web which is the image of π'. This curve J' is of order $4n(n-1)$ and has $2n(n-1)^2(4n-5)$ apparent double points. Its genus p' is, therefore, $p'=2n(n-1)(5n-8)+1$ which is also the genus of the plane section L' of L since L' and J' are in (1, 1) correspondence.

Knowing the order, class and genus of a plane section of L, the number of its nodes, cusps, bitangents δ', inflections κ' can be found by means of Plücker's equations. The numbers of its nodes and cusps are respectively the orders b and c of the nodal and cuspidal curves of L.

The genus D of L is the same as that found for J in §3, since L and J are in (1, 1) correspondence. L has no singularities other than a nodal and a cuspidal curve.

In the above, the values of the following characteristics of L have been obtained: N, n', a, a', b, b', c, c', δ, δ', κ, κ', D. From these, by use of the Cayley-Zeuthen equations,* the values of ρ, ρ', σ, σ', ϕ' result immediately and the remaining characteristics are found by the solution of sets of linear equations.

The characteristics of L obtained above are as follows:

$$N = 4n^2(n-1);$$

$$n' = 4(n-1)^3;$$

$$a' = a = 6n(n-1)^2;$$

$$b = 2n(n-1)(4n^4 - 4n^3 - 19n + 21);$$

$$b' = 2(n-1)^2(n-2)(4n^3 - 8n^2 + 8n - 25);$$

$$c = 2n(n-1)(11n-13);$$

$$c' = 30(n-1)^2(n-2);$$

$$\delta = (n-1)^2[18n^2(n-1)^2 - 59n + 74];$$

$$\delta' = n(n-1)[18n(n-1)^3 - 47n + 69];$$

$$\kappa = 12(n-1)^2(3n-4);$$

* Salmon, *Geometry of Three Dimensions*, 4th edition, 1882, pp. 596–600; Pascal, loc. cit., pp. 692–693.

$$\kappa' = 4n(n-1)(7n-11);$$
$$\sigma = 2(n-1)^2(19n-22);$$
$$\sigma' = 10n(n-1)(3n-5);$$
$$\rho = 4(n-1)^2(6n^4-6n^3-31n+34);$$
$$\rho' = 4n(n-1)[6(n-1)^4-25n+39];$$
$$\phi' = 2n(n-1)(73n-111);$$
$$r = 4(n-1)(26n^2-62n+37);$$
$$r' = 20(n-1)(6n^2-18n+13);$$
$$\beta = 8(n-1)(3n-4)(6n-7);$$
$$\beta' = 40(n-1)(2n-3)(3n-7);$$
$$q = 4(n-1)[6n^3(n-1)^2-49n^2+110n-62];$$
$$q' = 4(n-1)[6n(n-1)^4-55n^2+154n-105];$$
$$\gamma = 8(n-1)(11n^5-24n^4+13n^3-87n^2+207n-123);$$
$$\gamma' = 120(n-1)(n^5-6n^4+14n^3-25n^2+42n-31);$$
$$t = \frac{8}{3}(n-1)[4n^6(n-1)^2-57n^6+120n^4-63n^3+250n^2-574n+330];$$
$$t' = \frac{8}{3}(n-1)[4(n-1)^8-75n^5+438n^4-1002n^3+1498n^2-1937n+1256];$$
$$i = i' = 0;$$
$$D = \tfrac{1}{3}(2n-3)(4n-5)(4n-7).$$

6. **Loci of contacts and coincidences.** If the vertex P of a bundle of planes of (y) is on L, at the image point P_1 on J the surfaces of the corresponding net have a common tangent line l which lies in the tangent plane to J at P_1. This line l is tangent at P_1 to all the image curves of all the lines through P.

Of this net, the surfaces of a pencil corresponding to the planes of a pencil whose axis is tangent to L at P have contact at P_1. Since in the tangent plane to L at P, all the lines of the flat pencil with vertex at P are tangent to L at P, and since the axial pencils on the lines of this flat pencil include all the planes of the bundle on P, the net of surfaces which have contact with a line at P_1 is composed of a simply infinite linear system of pencils of surfaces such that the surfaces of each pencil have contact with each other at P_1, but do not have contact (except along the line) with any surface of another pencil. Of these surfaces, the one surface common to all the pencils, viz., the surface corresponding to the tangent plane to L at P, has a node at P_1.

The images on J of the cuspidal curve c and the nodal curve b of L are curves of orders c_1 and b_1 respectively.

If P lies on the cuspidal curve c of L, the surfaces of the corresponding net all osculate the curve c_1 at the image point P_1. Of this net, the surfaces of the pencil corresponding to the planes of the pencil whose axis is the tangent to c at P, all have stationary contact with each other at P_1.

If the surfaces of a net have a common osculating curve at a point P_1, three of the n^3 basis points of the net occur at P_1. Then the cuspidal curve of L is the locus of points P such that at each image point P_1 on J occurs

(1) the coincidence of three images of P;

(2) one stationary contact of the surfaces of a pencil of the web.

If P lies on the nodal curve b of L, the surfaces of the corresponding net all have two distinct contacts with the image curve b_1. One of these contacts is at P_1 on J, the image of P considered on one sheet of L through b; the other is at P_2 on J, the image of P considered on the other sheet of L through b. Of this net, the surfaces of the pencil corresponding to the planes through the tangent line to b at P all have contact with each other at P_1 and at P_2. Of this pencil, one surface corresponding to the tangent plane at P to one sheet of L has a node at P_1 and another surface corresponding to the tangent plane at P to the other sheet of L has a node at P_2.

If all the surfaces of a net have two distinct contacts with a curve at P_1 and P_2, two of the n^3 basis points coincide at P_1 and also two at P_2. Then the nodal curve of L is the locus of points P such that at each of the two distinct image points P_1 and P_2 of J occur

(1) two coincident images of P;

(2) one contact of the surfaces of a pencil of the web.

Since the points of J at which occur more than a simple coincidence, or at which occur a combination of simple coincidences, must lie also on the residual surface R, they must lie on the curves common to these two surfaces. The images of the nodal and cuspidal curves of L must, therefore, be the curves in which J and R intersect. The surfaces J and R intersect in a composite curve of order $16(n-1)^2(n^3-2)$.

The complete image of the nodal curve b of L is the intersection curve b_1 of J and R of order

$$b_1 = 4(n-1)(4n^4 - 4n^3 - 19n + 21)$$

counted twice and a residual curve b_2 on R of order

$$b_2 = 2(n-1)(n^3-4)(4n^4 - 4n^3 - 19n + 21),$$

which is the nodal curve of R. The complete images of b_1 and b_2 are b counted twice and n^3-4 times respectively. The curves b and b_1 are in (1, 2) point correspondence.

The complete image of the cuspidal curve c of L is the contact curve c_1 of J and R of order

$$c_1 = 2(n-1)(11n-13)$$

counted three times and a residual curve c_2 on R of order

$$c_2 = 2(n-1)(n^3-3)(11n-13),$$

which is the cuspidal curve of R. The complete images of c_1 and c_2 are c counted three times and n^3-3 times respectively. The curves c_1 and c are in (1, 1) correspondence.

From the above discussion of images of points of c, it results that the contact curve c_1 of J and R is the locus both of the three-point coincidences of the involution and of the stationary contacts of pencils of surfaces of the web. The n^3-3 residual images of each point of c lie on c_2.

Similarly, from the above discussion of images of points of b, it results that the intersection curve b_1 of J and R is the locus both of the pairs of simple coincidences of the involution and of the contacts at two distinct points of pencils of surfaces of the web. The n^3-4 residual images of each point of b lie on b_2.

At a point P on L, one of the γ intersections of the curves b and c which is a cusp on b, the two sheets of L through c are intersected by another sheet of L. The bundle of planes through P corresponds to a net of surfaces all of which (1) osculate the curve c_1 at P_3, the image of P considered as on the two sheets of L through c that have a common tangent plane, and (2) have contact with b_1 at P_2, the image of P considered as on the other sheet of L. Of the n^3 images of such a point P, three coincide at P_3 and two at P_2. The planes of the pencil whose axis is the cuspidal tangent to b at P correspond to surfaces of a pencil of the web which have stationary contact at P_3 and simple contact at P_2.

Therefore, the number of pencils of surfaces of the web that have two distinct contacts, one of which is stationary, and also the number of pairs of coincidences, each pair consisting of one three-point and one two-point coincidence, is

$$\gamma = 8(n-1)(11n^5-24n^4+13n^3-87n^2+207n-123).$$

The points P_3 lie at intersections of the curves b_1 and c_1 and the points P_2 lie on b_1. The curve c_2 passes through the points P_3 and b_2 passes through both the points P_3 and P_2. Then c_2 intersects J in γ points at the points P_3, and b_2 intersects J in 2γ points at the points P_3 and P_2.

The n^3-5 residual images of each of the γ points of L lie at n^3-5 intersections γ_2 of b_2 and c_2 which are cusps on b_2. The number $\gamma_2=(n^3-5)\gamma$.

Consider a point P on L at one of the β intersections of b and c which are cusps on c. At such a point, there is but one sheet of L. A plane through P intersects L in a curve with a triple point at P whose penultimate form consists of one node and two cusps. The bundle of planes on P corresponds to a net of surfaces all of which have four-point contact with c_1 at the image point P_4 on J. The pencil of planes whose axis is the cuspidal tangent to c at P corresponds to the pencil of surfaces which have tacnodal contact at P_4.

The number of pencils of surfaces of the web that have tacnodal contact and also the number of four-point coincidences of the involution is, therefore,

$$\beta = 8(n-1)(3n-4)(6n-7).$$

A four-point coincidence may be considered as (1) the union of two simple coincidences or (2) the addition of one image point to a triple coincidence. In case (1), the surfaces of a net which were all tangent to b_1 at two distinct points would all osculate b_1 at a single point. In case (2), the surfaces all of which osculated c_1 at one point would have four-point contact with c_1 at that point. Since both these possibilities must be accounted for, at the points P_4, b_1 osculates c_1. Since the three image points in the component triple coincidence of case (2) are symmetrical, the curve c_2 also osculates c_1 at the points P_4, and therefore c_2 has 3β intersections with J at these points. The curve b_2 does not pass through the points P_4.

The n^3-4 residual intersections of each of the β points of L lie at n^3-4 intersections β_2 of b_2 and c_2 which are cusps on c_2. The number $\beta_2=(n^3-4)\beta$.

A point P on L at a triple point of b corresponds to three distinct points P_2, P_2', P_2'' on b_1. To a bundle of planes through P corresponds a net of surfaces all of which have three distinct contacts with b_1 at P_2, P_2', P_2''. Of the n^3 basis points of the net, two coincide at each of these three points, that is, such a point P has three pairs of coincident image points.

A triple point P of b is also a triple point of L. Then at P there exists a single infinity of tangent planes to L enveloping the cubic tangent cone. At the three image points P_2, P_2', P_2'' of P, occur, therefore, nodes belonging to three singly infinite sets of distinct surfaces of the web corresponding to the tangent planes to the three respective sheets of L through P.

To the three tangent lines to b at P correspond three pencils of surfaces such that the surfaces of the pencils have two contacts at P_2, P_2'; P_2, P_2''; P_2', P_2'' respectively. One plane is common to each of the three pairs of axial pencils. These three planes correspond to three surfaces that have contact at all three image points.

Therefore the number of sets of surfaces, each set containing three surfaces that have three distinct contacts with each other, and also the number

of sets of three distinct simple coincidences in the involution, is the number of triple points of b,

$$t = \frac{8}{3}(n-1)[4n^6(n-1)^2 - 57n^5 + 120n^4 - 63n^3 + 250n^2 - 547n + 330].$$

The curve b_2 intersects b_1 at each of the three coincidences. Then at these points, b_2 intersects J in $3t$ points.

The residual images of the t points of L lie at $t_2 = (n^3-6)t$ triple points of b_2.

The tangent planes at the points β and γ of L are non-singular. To a tangent plane at a point β corresponds a surface with a node at the image point P_4. To the two tangent planes at a point γ correspond two distinct surfaces, one with a node at P_3 and the other with a node at P_2.

The curve c_2 intersects J in

$$3\beta + \gamma = 8(n-1)^2(n^3-3)(11n-13)$$

points, and the curve b_2 intersects J in

$$3t + 2\gamma = 8(n-1)^2(n^3-4)(4n^4 - 4n^3 - 19n + 21)$$

points, which occur as described above on the intersection curve b_1 and the contact curve c_1 of J and R.

7. **Loci of singularities of surfaces of the web.** As originally defined, the locus of the nodes of the surfaces of the web is J.

To a bitangent plane of L corresponds a surface of the web with two nodes. The two points of contact on L correspond respectively and uniquely to the two nodes of the image surface. The locus of the contacts of bitangent planes of L is the bitangential curve ρ' of order

$$\rho' = 4n(n-1)[6(n-1)^4 - 25n + 39].$$

The locus of the pairs of nodes, each pair of which belongs to one surface, is a curve ρ_1' on J, the image of ρ'. The complete image of ρ' is the curve ρ_1' of order

$$\rho_1' = 4(n-1)[6(n-1)^4 - 25n + 39]$$

counted twice and a residual curve ρ_2' on R of order

$$\rho_2' = 4(n-1)(n^3-2)[6(n-1)^4 - 25n + 39].$$

The curves ρ' and ρ_1' are in (1, 1) correspondence.

The number of intersections of ρ_1' and ρ_2' is

$$16(n-1)^2(n^3-2)[6(n-1)^4 - 25n + 39].$$

To a stationary plane of L corresponds a surface of the web with a binode, the binode being the image point on J of the point of contact on L. The locus

of the contacts of stationary tangent planes to L is the parabolic or spinodal curve σ' of order

$$\sigma' = 10n(n-1)(3n-5).$$

The complete image of σ' is the curve σ_1' on J of order

$$\sigma_1' = 10(n-1)(3n-5)$$

counted twice and a residual curve σ_2' on R of order

$$\sigma_2' = 10(n-1)(n^3-2)(3n-5).$$

The curve σ_1' on J is the locus of binodes of surfaces of the web.

The number of intersections of σ_1' and σ_2' is

$$40(n-1)^2(n^3-2)(3n-5).$$

The flecnodal curve ϕ' of order $2n(n-1)(73n-111)$ is the locus of contacts of flecnodal tangent planes of L. To a flecnodal tangent plane, corresponds a surface of the web with a node, belonging to a pencil of surfaces whose basis curve (the image of the inflectional tangent) osculates the tangent quadric cone of the node at its vertex. To a biflecnodal tangent plane whose contact occurs at a node of ϕ', corresponds a surface with a node. This surface is common to two pencils of surfaces both of whose basis curves osculate the tangent quadric cone of the node at its vertex. While these singularities account for two and three invariants respectively, actually the only singularity on the image surface in each case is a node. The nodes of the image surfaces of flecnodal and biflecnodal tangent planes of L all lie on a curve ϕ_1' of order $2(n-1)(73n-111)$, the image on J of ϕ'.

To a triple tangent plane of L corresponds a surface of the web with three nodes. The three contacts on L correspond respectively and uniquely to the three nodes of the image surface. The number of surfaces of the web each of which has three nodes is the number t' of triple tangent planes to L. This number is

$$t' = \frac{8}{3}(n-1)[4(n-1)^8 - 75n^5 + 438n^4 - 1002n^3 + 1498n^2 - 1937n + 1256].$$

These $3t'$ nodes lie at definite points, possibly nodes, of the curve ρ_1' on J.

To the nodo-cuspidal planes (tangent planes with two contacts one of which is stationary) of L correspond surfaces of the web each with one node and one binode. The node and the binode correspond respectively and uniquely to the simple and stationary contact. The nodo-cuspidal planes of L are the stationary planes γ' of the bitangential developable. Therefore the number of surfaces of the web that have both a node and a binode are

$$\gamma' = 120(n-1)(n^5 - 6n^4 + 14n^3 - 25n^2 + 42n - 31).$$

To the tacnodal tangent planes of L correspond surfaces of the web each of which has a special binode B_4 whose axis has four-point contact with the surface. The tacnodal contacts and the binodes B_4 are in (1, 1) correspondence. The tacnodal tangent planes of L are the stationary planes β' of the spinodal developable. Therefore the number of surfaces of the web with a B_4 is

$$\beta' = 40(n-1)(2n-3)(3n-7).$$

At the tacnodal points β' of L, the three curves ρ', σ' and ϕ' all have the tacnodal tangent as tangent and therefore all three have contact with each other at these points. The cuspidal curve c also passes through each point β', intersecting each of the above three curves at these points.

Since the cuspidal curve c passes through the β' tacnodal points of L, the β' binodes B_4 lie on the contact curve c_1 of J and R and therefore at the points B_4, three image points coincide. At these points ρ_1' and σ_1' are tangent and also the residual curves ρ_2' and σ_2' have contact with each other and with ρ_1' and σ_1'.

The intersections of σ_1' and σ_2' lie on the curves c_1 and b_1 common to J and R. The β' contacts count as $2\beta'$ of these intersections, leaving $40(n-1) \cdot [n^2(n-1)(3n-5)-2(n-2)(9n-13)]$ intersections of σ_1' and σ_2' that lie on b_1. This is likewise the number of intersections of b and σ' on L. To a point P of intersection of b and σ' correspond one point P_1 at which occur both a contact of surfaces of the net along b_1 and a binode, and another point P_2 at which occurs only a contact along b_1. The curves σ_1' and σ_2' intersect only at the points P_1 of b_1.

Of the intersections of ρ_1' and ρ_2', $2\beta'$ lie on c_1. The remaining $16(n-1) \cdot \{(n-1)(n^3-2)[6(n-1)^4-25n+39]-5(2n-3)(3n-7)\}$ intersections lie on b_1 and correspond uniquely to the intersections of b and ρ' on L. Associated with a point P of intersection of b and ρ' is a point P' on ρ', but not on b, which is the other contact of the bitangent plane π, one of whose contacts is at P. The image of P considered on the sheet of L to which π is tangent is a point P_1 of intersection of ρ_1' and ρ_2' on b_1. At P_1 occurs a node of the image surface f of π and also a contact of f and b_1. At P_2, the image of P considered on the other sheet of L is simply a contact of f (and the other surfaces of the net) with b_1. Also, f has a node at P_1', the image of P'.

In addition to β' contacts, the curves ρ' and σ' of L intersect at the γ' cuspidal points of the nodo-cuspidal planes. The binodes of the γ' surfaces of the web that have both a node and a binode therefore lie at γ' intersections of ρ_1' and σ_1' on J, and the associated nodes lie at fixed points on ρ_1'.

8. **The steinerian.** The jacobian surface J may also be defined as the locus of points P_1 whose polar planes with respect to all surfaces of the web are concurrent at points P_2. The points P_1 are nodes of surfaces of the web. The locus

of the associated points P_2 is a surface, the steinerian S of the web. J and S are thus in (1, 1) correspondence.

The branch-point surface L is the envelope of planes which are in (1, 1) correspondence with the nodes of surfaces of the web which, in turn, are in (1, 1) correspondence with the points of S. Then the tangent planes of L and the points of S are in (1, 1) correspondence. We have, therefore,

The branch-point surface L of the involution defined by the web and the steinerian S of the web are reciprocal surfaces.

All the characteristics of S may be obtained from the characteristics of L given at the end of §5 by merely interchanging the accented and unaccented symbols.

9. **Webs of quadrics.** In a web of quadrics, J and S coincide. L and J are not reciprocal surfaces, however.

Since three conditions are sufficient for a quadric to degenerate into two planes, the web contains a finite number of composite quadrics. This number is ten. The axes of these composite quadrics are ten lines on the quartic J. The images of these lines of J are ten conic tropes C' of L. L is of order 4 and class 16. Its reciprocal is a quartic surface with ten nodes.

The simplest method, although one not heretofore used, of obtaining the characteristics of the quadric web and its associated involution, is to determine the characteristics of L by deriving those of its reciprocal ten-nodal quartic. There results the following:

$$N = 16, \quad n' = 4, \quad a = a' = 12, \quad C' = 10,$$
$$\delta = 28, \quad \delta' = 22, \quad \kappa = \kappa' = 24, \quad \sigma = 32, \quad \rho = 80,$$
$$b = 60, \quad q = 40, \quad \gamma = 120, \quad t = 80,$$
$$c = 36, \quad r = 68, \quad \beta = 80,$$
$$\rho' = \sigma' = b' = c' = r' = \beta' = q' = \gamma' = t' = i = i' = 0.$$

Among the general formulas obtained for L in §5, all those defining the values of the unaccented characteristics yield the above values for $n=2$; but the correct values (all of which are zero) for $n=2$ can not be obtained from the general formulas for certain accented characteristics, namely, ρ', σ', r', β', q', γ', t'.

The cause of this apparent discrepancy is the existence of the ten conic tropes on L which result from the ten degenerate quadrics of the web. Three conditions are sufficient for an algebraic surface of order n to degenerate only when $n=2$, so that for $n>2$, no general web contains degenerate surfaces. In fact, the most general web of quadrics has the same properties as the polar web of a cubic surface, so that a general web of surfaces exists only for $n \geqq 3$.

WELLS COLLEGE,
AURORA, N. Y.

AN ITERATIVE PROCESS IN THE PROBLEM OF PLATEAU*

BY

T. RADÓ

Introduction

The problem of Plateau calls for a minimal surface bounded by a given curve. In analytic formulation, the problem requires the determination of three functions $x(u, v)$, $y(u, v)$, $z(u, v)$, subject to certain boundary conditions and such that

$$\text{(I)} \qquad x(u, v),\ \ y(u, v),\ \ z(u, v) \text{ are harmonic,}$$

and

$$\text{(II)} \qquad E = G,\ F = 0,$$

where

$$E = x_u^2 + y_u^2 + z_u^2, F = x_u x_v + y_u y_v + z_u z_v, G = x_v^2 + y_v^2 + z_v^2.$$

In my previous work† on this problem, I introduced the notion of *approximate solutions* of the problem by replacing the above *exact condition* (II) by the *approximate condition* that the two integrals

$$\iint (E^{1/2} - G^{1/2})^2 du dv \quad \text{and} \quad \iint |F|\, du dv$$

be small. The boundary conditions were also replaced by approximate conditions. I gave a direct construction for the approximate solutions; the exact solution was then obtained by a passage to the limit.

The main purpose of the present paper is to develop an *iterative process* which produces automatically a sequence of approximate solutions. The idea of the process (described and discussed in §2) is to comply with the above conditions (I) and (II) alternately. The process starts with an arbitrary harmonic surface bounded by the given curve and thus the result depends, in general, upon the choice of this initial harmonic surface. In this sense, the process contains an *arbitrary parameter*, namely the initial harmonic surface $\mathfrak{H}_0$, and this fact accounts for the great flexibility of the method.

* Presented to the Society, December 29, 1932; received by the editors February 16, 1933.

† *On Plateau's problem*, Annals of Mathematics, (2), vol. 31 (1930), pp. 457–469; *The problem of the least area and the problem of Plateau*, Mathematische Zeitschrift, vol. 32 (1930), pp. 763–796.

To illustrate this point, let us recall that in general the solution of the problem of Plateau is not unique.† The construction of the approximate solutions, used in my previous work, yielded a solution of the problem whose area was a minimum, and which consequently was not the general solution of the problem, since a minimal surface, in general, does not have a minimum area.‡ On the other hand, it is rather obvious that the iterative process yields *all the solutions* of the problem, including those with minimum area, if the parameter $\mathfrak{H}_0$ is chosen in all possible ways. For the sake of simplicity, this very trivial remark will be verified only for the case when the given boundary curve is a polygon. On the other hand, we shall see that the iterative process, if properly applied, can also be made to yield a *solution with a minimum area*.

After the approximate solutions have been constructed, a passage to the limit is necessary to obtain an exact solution. In my previous work, this passage to the limit has been carried out under the assumption that the given boundary curve bounds at least one continuous surface with a finite area. Replacing a lemma of Courant which I used in my proof by a lemma due to J. Douglas, it follows however that the restriction mentioned above can be dropped. Thus it follows that in order to obtain the solution for a general Jordan curve, it is sufficient to secure approximate solutions (which can be constructed directly); it is not necessary to solve the problem of Plateau first for some special class of curves.

Summing up, we have the following picture. Given a general Jordan curve, we can always construct approximate solutions of the problem of Plateau, and a passage to the limit yields then an exact solution. In case the given curve has a finite minimum area, the approximate solutions can be constructed in such a way as to obtain an exact solution with a minimum area. The construction of the approximate solutions can be based on an automatic iterative process, properly applied. In case the given curve is a polygon, the iterative process can be shown to yield all the solutions of the problem.

We conclude this introduction with a few remarks concerning the method used in this paper. The characteristic feature of the method is the essential use of *conformal mapping*. Indeed, the method is based on the following operation. Given a surface

$$S\colon x = x(u, v), \quad y = y(u, v), \quad z = z(u, v), \quad u^2 + v^2 \leqq 1;$$

first, change to isothermic parameters, that is to say, change to a representation

$$S\colon x = x^*(u, v),\ y = y^*(u, v),\ z = z^*(u, v),\ u^2 + v^2 \leqq 1,$$

† See also for literature the author's paper *Contributions to the theory of minimal surfaces*, Acta Szeged, vol. 6 (1932), pp. 1–20.

‡ See Radó, Acta Szeged, loc. cit.

such that $E^*=G^*$, $F^*=0$. Secondly, take the harmonic functions $\bar{x}(u, v)$, $\bar{y}(u, v)$, $\bar{z}(u, v)$ which coincide with $x^*(u, v)$, $y^*(u, v)$, $z^*(u, v)$ on $u^2+v^2=1$. Thus we derive from S a harmonic surface

$$\mathfrak{H}: x = \bar{x}(u, v),\ y = \bar{y}(u, v),\ z = \bar{z}(u, v),\ u^2 + v^2 \leqq 1.$$

This operation, leading from S to $\mathfrak{H}$, and the simple relations between S and $\mathfrak{H}$, were the essential tools which I used to construct the approximate solutions in my previous work†, and these same tools will be used also in the present paper in setting up the iterative process. In order to get from S to $\mathfrak{H}$, we have to use a conformal map of S, and to get around the difficulties arising in this connection, in my previous work I approximated S by polyhedrons and referred to the theorem, proved by H. A. Schwarz, that polyhedrons do admit of conformal maps (in a properly generalized sense). Since my first publications on this subject, very substantial progress has been achieved in the theory of the conformal mapping of general surfaces.‡ A theorem of McShane, concerned with conformal maps of saddle-shaped surfaces, permits us to deal directly with the surfaces which arise in the course of the iterative process and to present this process in a very compact and elegant way.

1. Preliminaries

1.1. In this section, the definitions, lemmas and theorems referred to in the sequel will be stated for the convenience of the reader.

A *Jordan arc*, in the xyz-space, is a one-to-one and continuous image of an interval $a \leqq t \leqq b$. A *Jordan curve* is the one-to-one and continuous image of the unit circle $u = \cos\theta$, $v = \sin\theta$.

If C_1 and C_2 are two Jordan arcs, then their *distance* $d(C_1, C_2)$, in the Fréchet sense, is defined as follows. Denote by τ a one-to-one and continuous correspondence between C_1 and C_2, and let $M(\tau)$ denote the maximum distance of corresponding points. The distance $d(C_1, C_2)$ is the greatest lower bound of $M(\tau)$, for all possible choices of τ. The distance of two Jordan curves is defined in the same manner. *Convergent sequences* are then defined in terms of the distance.

1.2. Given two Jordan arcs C_1 and C_2, suppose there is given a transfor-

† See second foot note on p. 869. This passage from S to $\mathfrak{H}$, and the simple inequalities concerning the relations between S and $\mathfrak{H}$, have been subsequently used by J. Douglas and E. J. McShane also. See J. Douglas, *Solution of the problem of Plateau*, these Transactions, vol. 33 (1931), pp. 263–321, and *The mapping theorem of Koebe and the problem of Plateau*, Journal of Mathematics and Physics of the Massachusetts Institute of Technology, vol. 10 (1931), pp. 106–130; E. J. McShane, these Transactions, vol. 35 (1933), pp. 716–733.

‡ McShane, loc. cit. Further important contributions to the theory of the conformal maps of general surfaces are contained in several as yet unpublished papers by C. B. Morrey, which the author has had the privilege to see in manuscript.

mation T which associates with every point of C_1 a definite point of C_2 (the existence of an inverse transformation is not required). T will be called a *monotonic transformation* of C_1 into a set on C_2 if the following conditions are satisfied. Whenever distinct points P_1, Q_1, R_1 on C_1 are such that Q_1 is on the sub-arc with end points P_1, R_1, then their images P_2, Q_2, R_2 on C_2 have the same relative positions; in case P_2 and R_2 coincide, it is required that Q_2 also coincide with them.

For two Jordan curves Γ_1^*, Γ_2^*, a monotonic transformation T† of Γ_1^* into a set on Γ_2^* is then defined as follows. There exists a triple of points A_1^*, B_1^*, C_1^* on Γ_1^* with distinct images A_2^*, B_2^*, C_2^*, such that the three non-overlapping arcs $A_1^*B_1^*$, $B_1^*C_1^*$, $C_1^*A_1^*$ of Γ_1^* are taken by T in a monotonic way into sets on the three non-overlapping arcs $A_2^*B_2^*$, $B_2^*C_2^*$, $C_2^*A_2^*$ of Γ_2^*. The notion of a continuous monotonic transformation is then self-explanatory.

1.3. The term *monotonic transformation* has been chosen to suggest that such transformations have a number of properties in common with monotonic functions. As a consequence, a simple and important selection theorem of Helly, concerned with sequences of monotonic functions, generalizes immediately to sequences of monotonic transformations in the following manner.‡

Let there be given a Jordan curve γ, and a sequence of Jordan curves Γ_n^* converging toward a Jordan curve Γ^* in the Fréchet sense. Given three distinct points a, b, c on γ, three distinct points A^*, B^*, C^* on Γ^*, and three distinct points A_n^*, B_n^*, C_n^* on Γ_n^*, such that $A_n^*\to A^*$, $B_n^*\to B^*$, $C_n^*\to C^*$. Consider any sequence of monotonic transformations T_n, such that T_n carries γ into a set on Γ_n^* and a, b, c into A_n^*, B_n^*, C_n^*. Then there exists a subsequence T_{n_k}, such that for every point p of γ the sequence of the points $P_{n_k}^*$, which correspond to p under T_{n_k}, converge toward a definite point P^* on Γ^*. The transformation which associates with p this limit point P^* is a monotonic transformation T of γ into a set on Γ^*. The limit transformation T carries a, b, c into A^*, B^*, C^*.

1.4. If a sequence of monotonic functions converges, in a closed interval, toward a continuous function, then the convergence necessarily is uniform.§ As observed by McShane in conversation with the author, this holds ob-

† J. Douglas, in his work on the problem of Plateau, speaks of proper and improper parametric representations, and obtains the necessary facts by an interpretation on the torus. The term monotonic transformation, used in my own work, calls attention to the analogy with monotonic functions; the necessary facts appear then as immediate consequences of this analogy.

‡ See references cited in second footnote on p. 869.

§ See H. E. Buchanan and T. H. Hildebrandt, *Note on the convergence of a sequence of functions of a certain type*, Annals of Mathematics, vol. 9 (1908), p. 123.

viously for sequences of monotonic transformations also. Thus we have the following corollary to the preceding selection theorem: if the limit transformation T is continuous, then T_{n_k} converges *uniformly* toward T, the meaning of this assertion being too obvious to be explained.

1.5. Suppose now we have a monotonic transformation T of the unit circle $u=\cos\theta$, $v=\sin\theta$ into a set on a Jordan curve Γ^*. Then T can be given by a set of equations

$$T: x = \xi(\theta),\ y = \eta(\theta),\ z = \zeta(\theta),$$

and, in analogy with monotonic functions, we have the following statements.

(a) If T is not a one-to-one transformation, then there exists an arc σ on $u^2+v^2=1$, such that $\xi(\theta)$, $\eta(\theta)$, $\zeta(\theta)$ all three reduce to constants on σ.

(b) For every θ_0, $\xi(\theta)$, $\eta(\theta)$, $\zeta(\theta)$ have definite one-sided limits ξ_0^+, ξ_0^-, η_0^+, η_0^-, ζ_0^+, ζ_0^-.

(c) If, for a certain θ_0, we have $\xi_0^+=\xi_0^-$, $\eta_0^+=\eta_0^-$, $\zeta_0^+=\zeta_0^-$, then $\xi(\theta)$, $\eta(\theta)$, $\zeta(\theta)$ are continuous at θ_0.

(d) The points of discontinuity of $\xi(\theta)$, $\eta(\theta)$, $\zeta(\theta)$ form a denumerable set.

1.6. A *continuous surface S, of the topological type of the circular disc*, is defined by a set of equations

$$S: x = x(u, v),\ y = y(u, v),\ z = z(u, v),\ (u, v) \text{ in } R, \tag{1}$$

where R is some Jordan region (that is, the set of points in and on a Jordan curve in the uv-plane), and $x(u, v)$, $y(u, v)$, $z(u, v)$ are continuous in R. Given then another such surface

$$\bar{S}: x = \bar{x}(u, v),\ y = \bar{y}(u, v),\ z = \bar{z}(u, v),\ (u, v) \text{ in } \bar{R}, \tag{2}$$

the *distance* $d(S, \bar{S})$, in the Fréchet sense, is defined as follows. Let τ denote a one-to-one and continuous correspondence between R and $\bar{R}$, and denote by (u, v), $(\bar{u}, \bar{v})$ a couple of points corresponding under τ. Denote by P the point of S corresponding to (u, v), by $\bar{P}$ the point of $\bar{S}$ corresponding to $(\bar{u}, \bar{v})$, and by $M(\tau)$ the maximum distance of P and $\bar{P}$, for all choices of the couple (u, v), $(\bar{u}, \bar{v})$. Then $d(S, \bar{S})$ is the greatest lower bound of $M(\tau)$, for all possible choices of τ.

If $d(S, \bar{S})=0$, then $\bar{S}$ is considered as identical to S, and (1) and (2) are considered as parametric representations of the same surface.

Convergent sequences of surfaces are defined in terms of the distance in an obvious manner.

A continuous surface S is bounded by a Jordan curve Γ^* if it admits of a representation (1) such that the boundary curve of R is taken in a topological way into Γ^*.

1.7. A continuous surface will be called a ***polyhedron*** and will be denoted by $\mathfrak{P}$, if it admits of a parametric representation

$$(3) \qquad \mathfrak{P}: x = x(u, v),\ y = y(u, v),\ z = z(u, v),\ (u, v) \text{ in } R,$$

with the following properties. R can be subdivided into a finite number of curvilinear triangles $\delta_1, \cdots, \delta_m$, every one of which is carried by (3) in a topological way into a non-degenerate plane rectilinear triangle in the xyz-space. These rectilinear triangles will be denoted by $\Delta_1, \cdots, \Delta_m$. It is furthermore required that the boundary curve of R be carried by (3) in a topological way into a simple closed polygon $\mathfrak{p}^*$, which will be called the boundary polygon of $\mathfrak{P}$. A representation (3) with these properties will be called a *typical representation* of $\mathfrak{P}$.

1.8. A fundamental theorem, proved already by H. A. Schwarz, asserts the existence of conformal maps of polyhedrons, in the following sense. Given a polyhedron $\mathfrak{P}$, there exists a representation which is typical in the sense of §1.7 and possesses the following additional properties.†

(a) The region R is the unit circle $u^2+v^2 \leqq 1$.

(b) The sides of the curvilinear triangles $\delta_1, \cdots, \delta_m$ are analytic arcs including their end points; none of these triangles has a zero angle.

(c) $x(u, v), y(u, v), z(u, v)$ are analytic in the interior of $\delta_1, \cdots, \delta_m$, and satisfy there the relations $E=G, F=0$.

(d) Three points A, B, C, given arbitrarily on $u^2+v^2=1$, are carried into three points A^*, B^*, C^* arbitrarily given on the boundary polygon $\mathfrak{p}^*$ of $\mathfrak{P}$.

1.9. The sum of the areas of the triangles $\Delta_1, \cdots, \Delta_m$, defined in §1.7, is the area $\mathfrak{A}(\mathfrak{P})$ of $\mathfrak{P}$. The *area* $\mathfrak{A}(S)$, in the Lebesgue sense, of a continuous surface S is then defined as follows. Consider a sequence of polyhedrons $\mathfrak{P}_n$ converging toward S in the Fréchet sense. $\mathfrak{A}(S)$ is then the greatest lower bound of lim inf $\mathfrak{A}(\mathfrak{P}_n)$ for all choices of the sequence $\mathfrak{P}_n$.‡

1.10. From the existence of conformal maps of polyhedrons McShane§ derived an important existence theorem for saddle-surfaces. To state this theorem, the following definitions are necessary.

A function $f(u, v)$, defined in $u^2+v^2 \leqq 1$, satisfies condition (C) if the following hold:

(a) $f(u, v)$ is continuous in $u^2+v^2 \leqq 1$.

---

† The reader may consult the beautiful book by Carathéodory, *Conformal Representation* (Cambridge University Press), Chapter VII.

‡ For a systematic presentation of the theory of the area and for literature, see the author's paper *Über das Flächenmass rektifizierbarer Flächen*, Mathematische Annalen, vol. 100 (1928), pp. 445–479. Quite recently, important contributions have been made to the theory by McShane (see for references McShane, loc. cit.) and by C. B. Morrey (in several as yet unpublished papers).

§ McShane, these Transactions, loc. cit.

(b) $f(u, v)$ is, for almost every value of u, an absolutely continuous function of v, and for almost every value of v an absolutely continuous function of u.

(c) The Dirichlet integral $\iint(f_u^2+f_v^2)$, taken over $u^2+v^2<1$, is finite.

1.11. A function $f(u, v)$, continuous in a Jordan region R, is *monotonic* there if for every domain D (connected open set) in R it is true that $m_b \leqq f(u, v) \leqq M_b$ in D, where m_b, M_b denote the minimum and maximum respectively of $f(u, v)$ on the boundary of D.

1.12. A continuous surface (1) is called a *saddle-surface* if $x(u, v)$, $y(u, v)$, $z(u, v)$ are monotonic. This property is independent of the parametric representation.†

1.13. The *theorem of McShane* reads then as follows. Given a continuous surface

$$(4) \qquad S: x = x(u, v),\ y = y(u, v),\ z = z(u, v),\ u^2 + v^2 \leqq 1,$$

suppose that

(a) S is a saddle-surface;

(b) $\mathfrak{A}(S)$ is finite;

(c) the equations (4) define a continuous monotonic transformation of $u^2+v^2=1$ into a Jordan curve Γ^*.

Then there exists a representation of S,

$$(5) \qquad S: x = \bar{x}(u, v),\ y = \bar{y}(u, v),\ z = \bar{z}(u, v),\ u^2 + v^2 \leqq 1,$$

which has the following properties.

(α) $\bar{x}(u, v)$, $\bar{y}(u, v)$, $\bar{z}(u, v)$ satisfy condition (C) of §1.10.

(β) $\bar{E}=\bar{G}$, $\bar{F}=0$ almost everywhere in $u^2+v^2<1$.

(γ) The equations (5) define again a continuous monotonic transformation of $u^2+v^2=1$ into Γ^*.

(δ) Three points A, B, C, given arbitrarily on $u^2+v^2=1$, are carried by (5) into three points A^*, B^*, C^* given arbitrarily on Γ^*.

(ϵ) $\mathfrak{A}(S)$ is given by the usual integral formula, which reduces, on account of (β), to

$$\mathfrak{A}(S) = \tfrac{1}{2} \iint_{u^2+v^2<1} (\bar{E} + \bar{G})\,du\,dv.$$

1.14. Given a Jordan curve Γ^* in the xyz-space, the *problem of Plateau* for Γ^* will be stated as follows. Determine three functions $x(u, v)$, $y(u, v)$, $z(u, v)$ with the following properties.

(a) $x(u, v)$, $y(u, v)$, $z(u, v)$ are continuous in $u^2+v^2\leqq 1$, harmonic in $u^2+v^2<1$, and

† McShane, these Transactions, loc. cit.

(b) satisfy in $u^2+v^2<1$ the equations $E=G$, $F=0$.

(c) The equations $x=x(u, v)$, $y=y(u, v)$, $z=z(u, v)$ carry $u^2+v^2=1$ in a topological way into Γ^*.

1.15. Given, in a domain D, three functions $x(u, v)$, $y(u, v)$, $z(u, v)$, we shall say that they form a *triple of conjugate harmonic functicns*, if they are harmonic and satisfy the equations $E=G$, $F=0$ in D. If one of the three functions, say $z(u, v)$, vanishes identically, then $x(u, v)$ and $y(u, v)$ are obviously conjugate harmonic functions in the sense used in theory of functions. This analogy between minimal surfaces on the one hand and analytic functions of a complex variable $w=u+iv$ on the other hand has always been of fundamental importance in the theory of minimal surfaces.

We shall need the following two lemmas.†

1.16. Suppose we have, in $u^2+v^2<1$, a triple of conjugate harmonic functions $x(u, v)$, $y(u, v)$, $z(u, v)$ which remain continuous on $u^2+v^2=1$, and all three reduce to constants on a certain arc of $u^2+v^2=1$. Then $x(u, v)$, $y(u, v)$, $z(u, v)$ reduce to constants identically.

1.17. Let there be given, in a sector $0<u^2+v^2<r^2$, $0<\arctan(v/u)<\alpha$, a triple of conjugate harmonic functions $x(u, v)$, $y(u, v)$, $z(u, v)$. Suppose that these functions remain continuous on $v=0$, $0<u<r$, and that $x(u, 0)$, $y(u, 0)$, $z(u, 0)$ approach definite finite limits x_0, y_0, z_0 for $u\to+0$. Then $x(u, v)\to x_0$, $y(u, v)\to y_0$, $z(u, v)\to z_0$ if $(u, v)\to(0, 0)$ in any subsector

$$0 < u^2 + v^2 < r^2, \quad 0 \leqq \arctan\frac{v}{u} \leqq \beta < \alpha.$$

1.18. Let there be given, on the unit circle $u=\cos\theta$, $v=\sin\theta$, three functions $\xi(\theta)$, $\eta(\theta)$, $\zeta(\theta)$. Suppose that $\xi(\theta)$, $\eta(\theta)$, $\zeta(\theta)$ are summable and that at a certain θ_0 they have definite finite one-sided limits ξ_0^+, ξ_0^-, η_0^+, η_0^-, ζ_0^+, ζ_0^-. Suppose that the harmonic functions $x(u, v)$, $y(u, v)$, $z(u, v)$, obtained by means of the Poisson integral formula by using $\xi(\theta)$, $\eta(\theta)$, $\zeta(\theta)$ as boundary functions, constitute a triple of conjugate harmonic functions in $u^2+v^2<1$. Then

$$\xi_0^+ = \xi_0^-, \quad \eta_0^+ = \eta_0^-, \quad \zeta_0^+ = \zeta_0^-.$$

This lemma, due to J. Douglas‡, can be obtained as an immediate consequence of the generalized Lindelöf theorem, stated in §1.16.§

1.19. Given, in the xyz-space, a Jordan curve Γ^*, we shall consider in the

† For proofs and literature concerning the lemmas in 1.16 and 1.17, which generalize classical theorems of Schwarz and of Lindelöf, see E. F. Beckenbach and T. Radó, *Subharmonic functions and minimal surfaces*, these Transactions, vol. 35 (1933), p. 648–661.

‡ Loc. cit. in first footnote on p. 871.

§ Beckenbach and Radó, loc. cit.

sequel an approximate form of the problem of Plateau so often that it is convenient to have a symbol for it. Given three distinct points A, B, C on $u^2+v^2=1$, three distinct points A^*, B^*, C^* on Γ^*, and an $\epsilon>0$. We shall then denote by $P(\Gamma^*; A, B, C, A^*, B^*, C^*; \epsilon)$ the following problem.† Determine three functions $x(u, v)$, $y(u, v)$, $z(u, v)$ with the following properties.

(a) $x(u, v)$, $y(u, v)$, $z(u, v)$ are continuous in $u^2+v^2\leqq 1$, harmonic in $u^2+v^2<1$, and

(b) satisfy the relations

$$\iint (E^{1/2} - G^{1/2})^2 \leqq \epsilon, \quad \iint |F| \leqq \epsilon,$$

the integrals being taken over $u^2+v^2<1$.

(c) The equations

$$x = x(u, v),\ y = y(u, v),\ z = z(u, v),\ u^2 + v^2 = 1, \tag{6}$$

define a continuous monotonic transformation (see §1.2) of $u^2+v^2=1$ into a (not prescribed) Jordan curve $\overline{\Gamma}^*$, such that the distance of Γ^* and $\overline{\Gamma}^*$ is $\leqq\epsilon$. Finally, A, B, C are taken by (6) into three distinct points $\overline{A}^*$, $\overline{B}^*$, $\overline{C}^*$ on $\overline{\Gamma}^*$, such that the distances $A^*\overline{A}^*$, $B^*\overline{B}^*$, $C^*\overline{C}^*$ are all three $\leqq\epsilon$.

1.20. It is important to observe that the problem $P(\Gamma^*; A, B, C, A^*, B^*, C^*; 0)$ is identical to the problem of Plateau, as stated in §1.14. Indeed, if $\epsilon=0$, then condition (b) in §1.19 reduces to $E=G$, $F=0$ in $u^2+v^2<1$, since E, F, G are continuous (even analytic). We have to see what happens on the boundary. If $\epsilon=0$, then condition (c) in §1.19 only requires that the equations (6) define a continuous and monotonic transformation T of $u^2+v^2=1$ into Γ^*. We must show that T is a one-to-one transformation. However, since $E=G$, $F=0$ has already been verified, the one-to-one character of T follows directly from §1.16.

2. The iterative process

2.1. Let there be given, in the xyz-space, a Jordan curve Γ^*. Take three distinct points A, B, C on $u^2+v^2=1$, and three distinct points A^*, B^*, C^* on Γ^*. Suppose there is given a triple of functions $x_0(u, v)$, $y_0(u, v)$, $z_0(u, v)$ with the following properties.

(α) $x_0(u, v)$, $y_0(u, v)$, $z_0(u, v)$ are continuous in $u^2+v^2\leqq 1$ and harmonic in $u^2+v^2<1$.

(β) The equations

$$x = x_0(u, v),\ y = y_0(u, v),\ z = z_0(u, v),\ u^2 + v^2 = 1,$$

† The approximate form of the problem of Plateau has been introduced in the author's papers in the Annals of Mathematics and Mathematische Zeitschrift, loc. cit.

define a continuous monotonic transformation of $u^2+v^2=1$ into Γ^*, such that A, B, C are taken into A^*, B^*, C^*.

(γ) The area of the surface

$$\mathfrak{H}_0: x = x_0(u, v),\ y = y_0(u, v),\ z = z_0(u, v),\ u^2 + v^2 \leqq 1, \tag{7}$$

is finite.

On account of condition (α), the area of $\mathfrak{H}_0$ is then given by

$$\mathfrak{A}(\mathfrak{H}_0) = \iint (E_0G_0 - F_0^2)^{1/2}$$

where E_0, F_0, G_0 are the first fundamental quantities relative to the representation (7).

2.2. The preceding assumptions being satisfied, the iterative process runs as follows. On account of conditions (α), (β), (γ) the theorem of McShane (§1.13) applies to $\mathfrak{H}_0$, and we have therefore a representation

$$\mathfrak{H}_0: x = \bar{x}_0(u, v),\ y = \bar{y}_0(u, v),\ z = \bar{z}_0(u, v),\ u^2 + v^2 \leqq 1 \tag{8}$$

which satisfies the following conditions.

($\bar{\alpha}$) $\bar{x}_0(u, v)$, $\bar{y}_0(u, v)$, $\bar{z}_0(u, v)$ satisfy condition (C) of §1.10 in $u^2+v^2\leqq 1$, and also satisfy there the relations $\overline{E}_0=\overline{G}_0$, $\overline{F}_0=0$ almost everywhere ($\overline{E}_0$, $\overline{F}_0$, $\overline{G}_0$ are the first fundamental quantities relative to the representation (8)).

($\bar{\beta}$) The equations (8) define, for $u^2+v^2=1$, a continuous monotonic transformation of $u^2+v^2=1$ into Γ^*, such that A, B, C are carried into A^*, B^*, C^*.

($\bar{\gamma}$) On account of ($\bar{\alpha}$), $\mathfrak{A}(\mathfrak{H}_0)$ is given by

$$\mathfrak{A}(\mathfrak{H}_0) = \tfrac{1}{2}\iint (\overline{E}_0 + \overline{G}_0). \tag{9}$$

Denote then by $x_1(u, v)$, $y_1(u, v)$, $z_1(u, v)$ the harmonic functions coinciding with $\bar{x}_0(u, v)$, $\bar{y}_0(u, v)$, $\bar{z}_0(u, v)$ on $u^2+v^2=1$, and define the surface $\mathfrak{H}_1$ by

$$\mathfrak{H}_1: x = x_1(u, v),\ y = y_1(u, v),\ z = z_1(u, v),\ u^2 + v^2 \leqq 1.$$

Let us first verify that $\mathfrak{H}_1$ has also a finite area. Since a harmonic function with given boundary values minimizes the Dirichlet integral† we have the inequalities

$$\iint (x_{1u}^2 + x_{1v}^2) \leqq \iint (\bar{x}_{0u}^2 + \bar{x}_{0v}^2), \quad \iint (y_{1u}^2 + y_{1v}^2) \leqq \iint (\bar{y}_{0u}^2 + \bar{y}_{0v}^2),$$

$$\iint (z_{1u}^2 + z_{1v}^2) \leqq \iint (\bar{z}_{0u}^2 + \bar{z}_{0v}^2),$$

† See Hurwitz-Courant, *Funktionentheorie* (Berlin, Springer, 1922), p. 335, Hilfsatz II. The proof, as given there, covers the present case on account of condition (C) stated in §1.10 (see McShane, loc. cit.).

and hence, by addition,

$$\iint (E_1 + G_1) \leqq \iint (\bar{E}_0 + \bar{G}_0), \tag{10}$$

the integrals being taken over $u^2+v^2<1$. Since

$$(E_1G_1 - F_1^2)^{1/2} \leqq E_1^{1/2}G_1^{1/2} \leqq \tfrac{1}{2}(E_1 + G_1),$$

it follows from (9) and from (10) that the integral of $(E_1G_1-F_1^2)^{1/2}$ taken over $u^2+v^2<1$, that is to say, the area of $\mathfrak{H}_1$, is finite. For further use we note the obvious inequalities

$$\begin{aligned} \mathfrak{A}(\mathfrak{H}_1) &= \iint (E_1G_1 - F_1^2)^{1/2} \leqq \iint E_1^{1/2}G_1^{1/2} \leqq \frac{1}{2}\iint (E_1 + G_1) \\ &\leqq \frac{1}{2}\iint (\bar{E}_0 + \bar{G}_0) = \mathfrak{A}(\mathfrak{H}_0) = \iint (E_0G_0 - F_0^2)^{1/2} \\ &\leqq \iint E_0^{1/2}G_0^{1/2} \leqq \frac{1}{2}\iint (E_0 + G_0). \end{aligned} \tag{11}$$

Since the area of $\mathfrak{H}_1$ is finite, we can repeat the procedure by which we derived $\mathfrak{H}_1$ from $\mathfrak{H}_0$. We obtain in this way a surface $\mathfrak{H}_2$, to which we can apply the same procedure, and so on indefinitely. We obtain in this way a sequence of surfaces

$$\mathfrak{H}_n\colon x = x_n(u, v),\ y = y_n(u, v),\ z = z_n(u, v),\ u^2 + v^2 \leqq 1, \tag{12}$$

where $x_n(u, v)$, $y_n(u, v)$, $z_n(u, v)$ satisfy the conditions (α), (β), (γ) listed in §2.1 (the points A, B, C, A^*, B^*, C^* are kept fixed).

2.3. We put

$$\epsilon_n = \max\left(\iint (E_n^{1/2} - G_n^{1/2})^2,\ \iint |F_n|\right).†$$

We assert that $\epsilon_n \to 0$. In other words, $x_n(u, v)$, $y_n(u, v)$, $z_n(u, v)$ constitute a solution of the problem $P(\Gamma^*; A,B,C,A^*, B^*, C^*; \epsilon_n)$, where $\epsilon_n \to 0$. Furthermore, we have $\mathfrak{A}(\mathfrak{H}_n) \leqq \mathfrak{A}(\mathfrak{H}_0)$, where $\mathfrak{H}_n$ and $\mathfrak{H}_0$ are given by (12) and (7) respectively. In other words, *every one of the harmonic surfaces* $\mathfrak{H}_n$ *has an area* $\leqq$ *the area of the initial harmonic surface* $\mathfrak{H}_0$.

2.4. To prove the preceding assertions, let us put

$$\mathfrak{A}_n = \mathfrak{A}(\mathfrak{H}_n) = \iint (E_nG_n - F_n^2)^{1/2},\quad \mathfrak{I}_n = \frac{1}{2}\iint (E_n + G_n).$$

---

† If a, b are two real numbers, then max (a, b) denotes the greater one of the two numbers (or their common value, if they are equal).

Then (11) yields the relations

$$\mathfrak{A}_1 \leqq \mathfrak{J}_1 \leqq \mathfrak{A}_0 \leqq \mathfrak{J}_0.$$

Since $\mathfrak{H}_{n+1}$ is derived from $\mathfrak{H}_n$ by the same procedure as $\mathfrak{H}_1$ from $\mathfrak{H}_0$, we have quite generally for $n=0, 1, 2, \cdots$ the relations

$$\mathfrak{A}_{n+1} \leqq \mathfrak{J}_{n+1} \leqq \mathfrak{A}_n \leqq \mathfrak{J}_n. \tag{13}$$

From (13) it follows that the sequences $\mathfrak{A}_0, \mathfrak{A}_1, \mathfrak{A}_2, \cdots$ and $\mathfrak{J}_0, \mathfrak{J}_1, \mathfrak{J}_2, \cdots$ are both descending. Since all the terms are $\geqq 0$, both sequences are convergent, and from (13) it follows then immediately that *both sequences converge toward the same limit*. Hence if we put $\sigma_n = \mathfrak{J}_n - \mathfrak{A}_n$, then

$$0 \leqq \sigma_n = \mathfrak{J}_n - \mathfrak{A}_n \to 0.$$

Since $\mathfrak{A}_0, \mathfrak{A}_1, \mathfrak{A}_2, \cdots$ is a descending sequence, we have also

$$\mathfrak{A}_n \leqq \mathfrak{A}_0. \tag{14}$$

Suppose in the sequel that $n \geqq 1$. From the relations

$$\mathfrak{A}_n = \iint (E_nG_n - F_n^2)^{1/2} \leqq \iint E_n^{1/2}G_n^{1/2} \leqq \frac{1}{2}\iint (E_n + G_n) \tag{15}$$

we infer

$$\begin{aligned} \frac{1}{2}\iint (E_n^{1/2} - G_n^{1/2})^2 &= \iint [\tfrac{1}{2}(E_n + G_n) - E_n^{1/2}G_n^{1/2}] \\ &= \mathfrak{J}_n - \iint E_n^{1/2}G_n^{1/2} \leqq \mathfrak{J}_n - \mathfrak{A}_n = \sigma_n. \end{aligned} \tag{16}$$

Furthermore, since

$$F_n| = [E_n^{1/2}G_n^{1/2} + (E_nG_n - F_n^2)^{1/2}]^{1/2}[E_n^{1/2}G_n^{1/2} - (E_nG_n - F_n^2)^{1/2}]^{1/2},$$

we infer, from (15), (16), and from the inequality of Schwarz, that

$$\begin{aligned} \left(\iint |F_n|\right)^2 &\leqq \iint [E_n^{1/2}G_n^{1/2} + (E_nG_n - F_n^2)^{1/2}] \times (\mathfrak{J}_n - \mathfrak{A}_n) \\ &\leqq (\mathfrak{J}_n + \mathfrak{A}_n)(\mathfrak{J}_n - \mathfrak{A}_n) = \sigma_n(\mathfrak{J}_n + \mathfrak{A}_n). \end{aligned} \tag{17}$$

On account of (13) we have, since $n \geqq 1$, the inequalities $\mathfrak{A}_n \leqq \mathfrak{J}_n \leqq \mathfrak{A}_0$. Hence, from (17),

$$\iint |F_n| \leqq 2^{1/2}\sigma_n^{1/2}\mathfrak{A}_0^{1/2}. \tag{18}$$

From (16) and (18) it follows that

$$\epsilon_n = \max\left(\iint (E_n^{1/2} - G_n^{1/2})^2, \iint |F_n|\right) \leqq \max\,(2\sigma_n, 2^{1/2}\sigma_n^{1/2}\mathfrak{A}_0^{1/2}).$$

Since $\mathfrak{A}_0$ is finite and $\sigma_n \to 0$, this proves that $\epsilon_n \to 0$. The last part of the statement in §2.3 has already been proved by (14).

2.5. We apply the preceding results first to the following situation. Let there be given, in the xyz-space, a simple closed polygon $\mathfrak{p}^*$, and let there also be given a polyhedron $\mathfrak{P}$ bounded by $\mathfrak{p}^*$. Given further three distinct points A, B, C on $u^2+v^2=1$, three distinct points A^*, B^*, C^* on $\mathfrak{p}^*$, and given an $\epsilon>0$.

Then there exists a solution

$$S: x = x(u, v),\ y = y(u, v),\ z = z(u, v),\ u^2 + v^2 \leqq 1,$$

of the problem $P(\mathfrak{p}^*; A, B, C, A^*, B^*, C^*; \epsilon)$ *such that* $\mathfrak{A}(S) \leqq \mathfrak{A}(\mathfrak{P})$. In particular (disregarding the condition concerned with the areas), *the problem* $P(\mathfrak{p}^*; A, B, C, A^*, B^*, C^*; \epsilon)$ *has a solution for every choice of* $A, B, C, A^*, B^*, C^*, \epsilon$.

To see this, let

$$\mathfrak{P}: x = \tilde{x}(u, v),\ y = \tilde{y}(u, v),\ z = \tilde{z}(u, v),\ u^2 + v^2 \leqq 1,$$

be an isothermic representation of $\mathfrak{P}$, such that A, B, C are taken into A^*, B^*, C^*. Let $x_0(u, v), y_0(u, v), z_0(u, v)$ be the harmonic functions coinciding with $\tilde{x}(u, v), \tilde{y}(u, v), \tilde{z}(u, v)$ on $u^2+v^2=1$, and denote by $\mathfrak{H}_0$ the surface

$$\mathfrak{H}_0: x = x_0(u, v),\ y = y_0(u, v),\ z = z_0(u, v),\ u^2 + v^2 \leqq 1.$$

On account of the minimizing property of harmonic functions we have again

$$\iint (E_0 + G_0) \leqq \iint (\tilde{E} + G). \tag{19}$$

From the inequalities $(E_0G_0 - F_0^2)^{1/2} \leqq \iint E_0^{1/2}G_0^{1/2} \leqq \iint (E_0+G_0)/2$, from the equations $\tilde{E}=\tilde{G}$, $\tilde{F}=0$ and from (19) we infer that

$$\mathfrak{A}(\mathfrak{H}_0) \leqq \frac{1}{2}\iint (E_0 + G_0) \leqq \frac{1}{2}\iint (\tilde{E} + \tilde{G}) = \mathfrak{A}(\mathfrak{P}). \tag{20}$$

Thus the area of $\mathfrak{H}_0$ is finite. Hence we can start up the iterative process, beginning with $\mathfrak{H}_0$. There results a sequence of solutions

$$\mathfrak{H}_n: x = x_n(u, v),\ y = y_n(u, v),\ z = z_n(u, v),\ u^2 + v^2 \leqq 1,$$

of the problems $P(\mathfrak{p}^*;\ A, B, C, A^*, B^*, C^*;\ \epsilon_n)$, where $\epsilon_n \to 0$ and $\mathfrak{A}(\mathfrak{H}_n) \leqq \mathfrak{A}(\mathfrak{H}_0)$. Hence, on account of (20), $\mathfrak{A}(\mathfrak{H}_n) \leqq \mathfrak{A}(\mathfrak{P})$ for every n, while ϵ_n will be $<\epsilon$ for n large enough. Thus, for n large enough, $\mathfrak{H}_n$ can be used as the

surface S whose existence has been asserted at the beginning of the present section.

2.6. Consider next a Jordan curve Γ^* in the xyz-space. Given three distinct points A, B, C on $u^2+v^2=1$, three distinct points A^*, B^*, C^* on Γ^*, and an $\epsilon>0$. *Then there exists a solution of the problem* $P(\Gamma^*; A, B, C, A^*, B^*, C^*; \epsilon)$.

To see this, observe first that we can approximate Γ^*, in the sense of Fréchet, by simple closed polygons. Hence we have a polygon $\mathfrak{p}^*$ such that $d(\mathfrak{p}^*, \Gamma^*)<\epsilon/2$. From the definition of the distance of two curves it follows then that we have on $\mathfrak{p}^*$ three distinct points $\overline{A}^*, \overline{B}^*, \overline{C}^*$ such that the distances $A^*\overline{A}^*, B^*\overline{B}^*, C^*\overline{C}^*$ are all three $<\epsilon/2$. On account of §2.5, the problem $P(\mathfrak{p}^*; A, B, C, \overline{A}^*, \overline{B}^*, \overline{C}^*; \epsilon/2)$ has a solution, and this solution is clearly a solution of the problem $P(\Gamma^*; A, B, C, A^*, B^*, C^*; \epsilon)$, as follows immediately from the definition of this problem (see §1.19)

2.7. Consider finally a Jordan curve Γ^* which bounds some continuous surface, of the topological type of the circular disc, with a finite area. Then the greatest lower bound $\mathfrak{a}(\Gamma^*)$ of the areas of all such surfaces (bounded by Γ^*) is finite. Given three distinct points A, B, C on $u^2+v^2=1$, three distinct points A^*, B^*, C^* on Γ^*, and an $\epsilon>0$. *Then there exists a solution*

$$S: x = x(u, v),\ y = y(u, v),\ z = z(u, v),\ u^2 + v^2 \leqq 1,$$

of the problem $P(\Gamma^*; A, B, C, A^*, B^*, C^*; \epsilon)$, *such that* $\mathfrak{A}(S)\leqq \mathfrak{a}(\Gamma^*)+\epsilon$.

To see this, observe first that we have, on account of the definition of $\mathfrak{a}(\Gamma^*)$, a continuous surface S_0, bounded by Γ^*, such that

$$\mathfrak{A}(S_0) < \mathfrak{a}(\Gamma^*) + \frac{\epsilon}{2}\cdot \tag{21}$$

On account of the definition of the area, we have then a polyhedron $\mathfrak{P}$, such that

$$d(S_0, \mathfrak{P}) < \frac{\epsilon}{2}, \quad \mathfrak{A}(\mathfrak{P}) < \mathfrak{A}(S_0) + \frac{\epsilon}{2}\cdot \tag{22}$$

On account of the definition of $d(S_0, \mathfrak{P})$, the distance of the boundary polygon $\mathfrak{p}^*$ of $\mathfrak{P}$ and of Γ^* is $<\epsilon/2$. Hence we have three distinct points $\overline{A}^*, \overline{B}^*, \overline{C}^*$ on $\mathfrak{p}^*$ such that the distances $A^*\overline{A}^*, B^*\overline{B}^*, C^*\overline{C}^*$ are all three $<\epsilon/2$. On account of §2.5, we have then a solution S of the problem $P(\mathfrak{p}^*; A, B, C, A^*, B^*, C^*; \epsilon/2)$ such that $\mathfrak{A}(S)\leqq\mathfrak{A}(\mathfrak{P})$. This S is then clearly a solution of the problem $P(\Gamma^*; A, B, C, A^*, B^*, C^*; \epsilon)$ and it also satisfies the inequality $\mathfrak{A}(S)<\mathfrak{a}(\Gamma^*)+\epsilon$, on account of (21) and (22).

3. A SELECTION THEOREM

3.1. Let Γ^* be a Jordan curve in the xyz-space. Choose three distinct points A, B, C on $u^2+v^2=1$ and three distinct points A^*, B^*, C^* on Γ^*. Suppose we have a sequence $\epsilon_n>0$, $\epsilon_n\to 0$, such that the problem $P(\Gamma^*; A, B, C, A^*, B^*, C^*; \epsilon_n)$ has a solution

$$(23)\qquad S_n\colon x = x_n(u, v),\ y = y_n(u, v),\ z = z_n(u, v),\ u^2 + v^2 \leqq 1,$$

for $n=1, 2, 3, \cdots$.

Then the sequence (23) *contains a subsequence* S_{n_k} *such that* $x_{n_k}(u, v)$, $y_{n_k}(u, v)$, $z_{n_k}(u, v)$ *converge uniformly in* $u^2+v^2\leqq 1$. *The limit surface*

$$S\colon x = x(u, v),\ y = y(u, v),\ z = z(u, v),\ u^2 + v^2 \leqq 1,$$

is a solution of the problem $P(\Gamma^*; A, B, C, A^*, B^*, C^*; 0)$. *If* $\liminf \mathfrak{A}(S_{n_k})$ *is finite, then* $\mathfrak{A}(S)$ *is also finite and satisfies the inequality*

$$\mathfrak{A}(S) \leqq \liminf \mathfrak{A}(S_{n_k}).$$

3.2. This theorem includes, as special cases, a number of selection theorems used previously in the literature.† The proof of the theorem runs as follows. If we put, for the sake of clarity,

$$\xi_n(\theta) = x_n(\cos\theta, \sin\theta),\ \eta_n(\theta) = y_n(\cos\theta, \sin\theta),\ \zeta_n(\theta) = z_n(\cos\theta, \sin\theta),$$

then the equations

$$x = \xi_n(\theta),\ y = \eta_n(\theta),\ z = \zeta_n(\theta)$$

define by assumption a sequence of monotonic transformations, and we can apply to this sequence the generalization of the selection theorem of Helly (see §1.3). There exists therefore a subsequence

$$x = \xi_{n_k}(\theta),\ y = \eta_{n_k}(\theta),\ z = \zeta_{n_k}(\theta)$$

which converges everywhere on the unit circle $u=\cos\theta$, $v=\sin\theta$. The limit functions $\xi(\theta)$, $\eta(\theta)$, $\zeta(\theta)$ define a monotonic transformation

$$(24)\qquad T\colon x = \xi(\theta),\ y = \eta(\theta),\ z = \zeta(\theta)$$

---

† R. Garnier, *Sur le problème de Plateau*, Annales Scientifiques de l'Ecole Normale, vol. 45 (1928), pp. 53–144; T. Radó, *Some remarks on the problem of Plateau*, Proceedings of the National Academy of Sciences, vol. 16 (1930), pp. 242–248, and Annals of Mathematics and Mathematische Zeitschrift, loc. cit.; J. Douglas, loc. cit. in the first foot note on p. 871.

In the development of my own work, I was guided by the analogy with a theorem, stated by Carathéodory, concerning the conformal maps of variable plane Jordan regions (see R. Courant, *Über eine Eigenschaft der Abbildungsfunktionen bei konformen Abbildung*, Göttinger Nachrichten, 1914 and a notice in 1922; T. Radó, *Sur la représentation conforme de domaines variables*, Acta Szeged, vol. 1 (1923)).

surface S whose existence has been asserted at the beginning of the present section.

2.6. Consider next a Jordan curve Γ^* in the xyz-space. Given three distinct points A, B, C on $u^2+v^2=1$, three distinct points A^*, B^*, C^* on Γ^*, and an $\epsilon>0$. *Then there exists a solution of the problem* $P(\Gamma^*; A, B, C, A^*, B^*, C^*; \epsilon)$.

To see this, observe first that we can approximate Γ^*, in the sense of Fréchet, by simple closed polygons. Hence we have a polygon $\mathfrak{p}^*$ such that $d(\mathfrak{p}^*, \Gamma^*)<\epsilon/2$. From the definition of the distance of two curves it follows then that we have on $\mathfrak{p}^*$ three distinct points $\overline{A}^*, \overline{B}^*, \overline{C}^*$ such that the distances $A^*\overline{A}^*$, $B^*\overline{B}^*$, $C^*\overline{C}^*$ are all three $<\epsilon/2$. On account of §2.5, the problem $P(\mathfrak{p}^*; A, B, C, \overline{A}^*, \overline{B}^*, \overline{C}^*; \epsilon/2)$ has a solution, and this solution is clearly a solution of the problem $P(\Gamma^*; A, B, C, A^*, B^*, C^*; \epsilon)$, as follows immediately from the definition of this problem (see §1.19)

2.7. Consider finally a Jordan curve Γ^* which bounds some continuous surface, of the topological type of the circular disc, with a finite area. Then the greatest lower bound $\mathfrak{a}(\Gamma^*)$ of the areas of all such surfaces (bounded by Γ^*) is finite. Given three distinct points A, B, C on $u^2+v^2=1$, three distinct points A^*, B^*, C^* on Γ^*, and an $\epsilon>0$. *Then there exists a solution*

$$S: x = x(u, v),\ y = y(u, v),\ z = z(u, v),\ u^2 + v^2 \leqq 1,$$

of the problem $P(\Gamma^*; A, B, C, A^*, B^*, C^*; \epsilon)$, *such that* $\mathfrak{A}(S)\leqq \mathfrak{a}(\Gamma^*)+\epsilon$.

To see this, observe first that we have, on account of the definition of $\mathfrak{a}(\Gamma^*)$, a continuous surface S_0, bounded by Γ^*, such that

$$\mathfrak{A}(S_0) < \mathfrak{a}(\Gamma^*) + \frac{\epsilon}{2} \cdot \tag{21}$$

On account of the definition of the area, we have then a polyhedron $\mathfrak{P}$, such that

$$d(S_0, \mathfrak{P}) < \frac{\epsilon}{2}, \quad \mathfrak{A}(\mathfrak{P}) < \mathfrak{A}(S_0) + \frac{\epsilon}{2} \cdot \tag{22}$$

On account of the definition of $d(S_0, \mathfrak{P})$, the distance of the boundary polygon $\mathfrak{p}^*$ of $\mathfrak{P}$ and of Γ^* is $<\epsilon/2$. Hence we have three distinct points $\overline{A}^*, \overline{B}^*, \overline{C}^*$ on $\mathfrak{p}^*$ such that the distances $A^*\overline{A}^*$, $B^*\overline{B}^*$, $C^*\overline{C}^*$ are all three $<\epsilon/2$. On account of §2.5, we have then a solution S of the problem $P(\mathfrak{p}^*; A, B, C, A^*, B^*, C^*; \epsilon/2)$ such that $\mathfrak{A}(S)\leqq \mathfrak{A}(\mathfrak{P})$. This S is then clearly a solution of the problem $P(\Gamma^*; A, B, C, A^*, B^*, C^*; \epsilon)$ and it also satisfies the inequality $\mathfrak{A}(S)<\mathfrak{a}(\Gamma^*)+\epsilon$, on account of (21) and (22).

3. A selection theorem

3.1. Let Γ^* be a Jordan curve in the xyz-space. Choose three distinct points A, B, C on $u^2+v^2=1$ and three distinct points A^*, B^*, C^* on Γ^*. Suppose we have a sequence $\epsilon_n>0$, $\epsilon_n\to 0$, such that the problem $P(\Gamma^*; A, B, C, A^*, B^*, C^*; \epsilon_n)$ has a solution

$$(23)\qquad S_n: x = x_n(u, v),\ y = y_n(u, v),\ z = z_n(u, v),\ u^2 + v^2 \leqq 1,$$

for $n=1, 2, 3, \cdots$.

Then the sequence (23) *contains a subsequence S_{n_k} such that $x_{n_k}(u, v)$, $y_{n_k}(u, v)$, $z_{n_k}(u, v)$ converge uniformly in $u^2+v^2\leqq 1$. The limit surface*

$$S: x = x(u, v),\ y = y(u, v),\ z = z(u, v),\ u^2 + v^2 \leqq 1,$$

is a solution of the problem $P(\Gamma^*; A, B, C, A^*, B^*, C^*; 0)$. If $\liminf \mathfrak{A}(S_{n_k})$ *is finite, then $\mathfrak{A}(S)$ is also finite and satisfies the inequality*

$$\mathfrak{A}(S) \leqq \liminf \mathfrak{A}(S_{n_k}).$$

3.2. This theorem includes, as special cases, a number of selection theorems used previously in the literature.† The proof of the theorem runs as follows. If we put, for the sake of clarity,

$$\xi_n(\theta) = x_n(\cos\theta, \sin\theta),\ \eta_n(\theta) = y_n(\cos\theta, \sin\theta),\ \zeta_n(\theta) = z_n(\cos\theta, \sin\theta),$$

then the equations

$$x = \xi_n(\theta),\ y = \eta_n(\theta),\ z = \zeta_n(\theta)$$

define by assumption a sequence of monotonic transformations, and we can apply to this sequence the generalization of the selection theorem of Helly (see §1.3). There exists therefore a subsequence

$$x = \xi_{n_k}(\theta),\ y = \eta_{n_k}(\theta),\ z = \zeta_{n_k}(\theta)$$

which converges everywhere on the unit circle $u=\cos\theta$, $v=\sin\theta$. The limit functions $\xi(\theta)$, $\eta(\theta)$, $\zeta(\theta)$ define a monotonic transformation

$$(24)\qquad T: x = \xi(\theta),\ y = \eta(\theta),\ z = \zeta(\theta)$$

---

† R. Garnier, *Sur le problème de Plateau*, Annales Scientifiques de l'Ecole Normale, vol. 45 (1928), pp. 53–144; T. Radó, *Some remarks on the problem of Plateau*, Proceedings of the National Academy of Sciences, vol. 16 (1930), pp. 242–248, and Annals of Mathematics and Mathematische Zeitschrift, loc. cit.; J. Douglas, loc. cit. in the first foot note on p. 871.

In the development of my own work, I was guided by the analogy with a theorem, stated by Carathéodory, concerning the conformal maps of variable plane Jordan regions (see R. Courant, *Über eine Eigenschaft der Abbildungsfunktionen bei konformen Abbildung*, Göttinger Nachrichten, 1914 and a notice in 1922; T. Radó, *Sur la représentation conforme de domaines variables*, Acta Szeged, vol. 1 (1923)).

of the unit circle $u=\cos\theta$, $y=\sin\theta$ into a set on Γ^*, such that A, B, C are taken into A^*, B^*, C^*. Denote by $x(u, v)$, $y(u, v)$, $z(u, v)$ the harmonic functions obtained by means of the Poisson integral formula, using $\xi(\theta)$, $\eta(\theta)$, $\zeta(\theta)$ as boundary functions.† We are going to discuss the surface

$$S: x = x(u, v),\ y = y(u, v),\ z = z(u, v), \tag{25}$$

which is defined, for the time being, only in $u^2+v^2<1$.

3.3. It follows from the Poisson integral formula that the harmonic functions $x_{n_k}(u, v)$, $y_{n_k}(u, v)$, $z_{n_k}(u, v)$ and all of their partial derivatives converge, in $u^2+v^2<1$, toward $x(u, v)$, $y(u, v)$, $z(u, v)$ and their corresponding partial derivatives, the convergence being uniform in every concentric circle $u^2+v^2\leqq r^2<1$. Consequently we have

$$\iint_{(r)} (E_{n_k}^{1/2} - G_{n_k}^{1/2})^2 \to \iint_{(r)} (E^{1/2} - G^{1/2})^2, \tag{26}$$

$$\iint_{(r)} |F_{n_k}| \to \iint_{(r)} |F|, \tag{27}$$

$$\iint_{(r)} (E_{n_k}G_{n_k} - F_{n_k}^2)^{1/2} \to \iint_{(r)} (EG - F^2)^{1/2}, \tag{28}$$

for every r such that $0<r<1$, the symbol (r) indicating that the integrals are taken over $u^2+v^2<r^2$. We also have

$$\iint_{(r)} (E_{n_k}^{1/2} - G_{n_k}^{1/2})^2 \leqq \iint_{(1)} (E_{n_k}^{1/2} - G_{n_k}^{1/2})^2 \leqq \epsilon_{n_k}, \tag{29}$$

$$\iint_{(r)} |F_{n_k}| \leqq \iint_{(1)} |F_{n_k}| \leqq \epsilon_{n_k}, \tag{30}$$

$$\iint_{(r)} (E_{n_k}G_{n_k} - F_{n_k}^2)^{1/2} \leqq \iint_{(1)} (E_{n_k}G_{n_k} - F_{n_k}^2)^{1/2} = \mathfrak{A}(S_{n_k}), \tag{31}$$

where the last line is, of course, to be considered only if $\mathfrak{A}(S_{n_k})$ is finite. From (26) to (31) it follows then, on account of $\epsilon_{n_k}\to 0$, that

$$\iint_{(r)} (E^{1/2} - G^{1/2})^2 = 0, \quad \iint_{(r)} |F| = 0, \tag{32}$$

$$\iint_{(r)} (EG - F^2)^{1/2} \leqq \liminf \mathfrak{A}(S_{n_k}). \tag{33}$$

From (33) it follows, in case $\liminf \mathfrak{A}(S_{n_k})$ happens to be finite, for $r\to 1$ that

† On account of §1.5, (d), the functions $\xi(\theta)$, $\eta(\theta)$, $\zeta(\theta)$ are integrable even in the Riemann sense.

$$\mathfrak{A}(S) = \iint_{(1)} (EG - F^2)^{1/2} \leqq \liminf \mathfrak{A}(S_{n_k}). \tag{34}$$

From (32) it follows, since E, F, G are continuous (and even analytic), that $E=G$, $F=0$ for $u^2+v^2<r^2$. Since r is arbitrary, it follows that

$$E = G, F = 0 \text{ in } u^2 + v^2 < 1. \tag{35}$$

3.4. Since (24) is a monotonic transformation, $\xi(\theta)$, $\eta(\theta)$, $\zeta(\theta)$ have the properties listed in §1.5. On account of (35), it follows therefore from the lemma of Douglas (see §1.18), in connection with §1.5, (c), that $\xi(\theta)$, $\eta(\theta)$, $\zeta(\theta)$ are continuous on the whole unit circle $u=\cos\theta$, $v=\sin\theta$. Consequently the harmonic functions $x(u, v)$, $y(u, v)$, $z(u, v)$ remain continuous on $u^2+v^2=1$ (that is to say, they are continuous in $u^2+v^2\leqq 1$), and we have

$$x(u, v) = \xi(\theta),\ y(u, v) = \eta(\theta),\ z(u, v) = \zeta(\theta) \text{ on } u^2 + v^2 = 1. \tag{36}$$

From this it follows that the equations (24) carry $u^2+v^2=1$ in a one-to-one way into Γ^*. Otherwise there would exist (see §1.5 (a)) an arc σ of $u^2+v^2=1$, such that $\xi(\theta)$, $\eta(\theta)$, $\zeta(\theta)$ all three reduce to constants on σ. On account of (36) and (35) it would then follow from §1.16 that $x(u, v)$, $y(u, v)$, $z(u, v)$ and consequently (cf. (36)) also $\xi(\theta)$, $\eta(\theta)$, $\zeta(\theta)$ reduce to constants identically. This contradicts the fact that the equations (24) carry the points A, B, C into three distinct points A^*, B^*, C^*.

3.5. Since it has been established in §3.4 that $\xi(\theta)$, $\eta(\theta)$, $\zeta(\theta)$ are continuous, it follows from §1.4 that $\xi_{n_k}(\theta)$, $\eta_{n_k}(\theta)$, $\zeta_{n_k}(\theta)$ converge uniformly toward $\xi(\theta)$, $\eta(\theta)$, $\zeta(\theta)$. From the principle of maximum it follows then that the harmonic functions $x_{n_k}(u, v)$, $y_{n_k}(u, v)$, $z_{n_k}(u, v)$ converge uniformly toward $x(u, v)$, $y(u, v)$, $z(u, v)$ in $u^2+v^2\leqq 1$. This completes the proof of the theorem stated in §3.1.

4. Applications

4.1. Let there be given, in the xyz-space, a Jordan curve Γ^*. Given three distinct points A, B, C on $u^2+v^2=1$, and three distinct points A^*, B^*, C^* on Γ^*. Then the problem $P(\Gamma^*; A, B, C, A^*, B^*, C^*; 1/n)$ is solvable for every positive integer n (see §2.6). Denote by

$$S_n\colon x = x_n(u, v),\ y = y_n(u, v),\ z = z_n(u, v),\ u^2 + v^2 \leqq 1, \tag{37}$$

a solution of this problem. On account of the selection theorem (see §3.1), a properly chosen subsequence of (37) will converge uniformly in $u^2+v^2\leqq 1$, and the limit surface solves the problem $P(\Gamma^*; A, B, C, A^*, B^*, C^*; 0)$, that is to say, the problem of Plateau for Γ^* (see §1.20). In other words: *the problem of Plateau is solvable for every Jordan curve.*†

† This result was first obtained by J. Douglas, loc. cit. in the first foot note on p. 871.

4.2. Suppose now that Γ^* bounds some continuous surface, of the topological type of the circular disc, with a finite area. Then the greatest lower bound $\mathfrak{a}(\Gamma^*)$ of the areas of all continuous surfaces, of the topological type of the circular disc and bounded by Γ^*, is finite. The problem $P(\Gamma^*; A, B, C, A^*, B^*, C^*; 1/n)$ has then (see §2.7) a solution

$$S_n: x = x_n(u, v),\ y = y_n(u, v),\ z = z_n(u, v),\ u^2 + v^2 \leqq 1,$$

such that

$$\mathfrak{A}(S_n) < \mathfrak{a}(\Gamma^*) + 1/n. \tag{38}$$

On account of the selection theorem of §3.1, a properly chosen subsequence S_{n_k} will converge, uniformly in $u^2+v^2 \leqq 1$, toward a solution S of the problem $P(\Gamma^*; A, B, C, A^*, B^*, C^*; 0)$, such that

$$\mathfrak{A}(S) \leqq \liminf \mathfrak{A}(S_{n_k}). \tag{39}$$

From (38) and (39) it follows that $\mathfrak{A}(S) \leqq \mathfrak{a}(\Gamma^*)$. On the other hand we have also $\mathfrak{A}(S) \geqq \mathfrak{a}(\Gamma^*)$, since S is a continuous surface, of the type of the circular disc, bounded by Γ^*. Hence $\mathfrak{A}(S) = \mathfrak{a}(\Gamma^*)$. In other words: *if S bounds some continuous surface (of the topological type of the circular disc) with a finite area, then there exists a minimal surface, bounded by Γ^*, whose area is a minimum with respect to all continuous surfaces, of the topological type of the circular disc, bounded by Γ^*.*†

4.3. Suppose now that Γ^* is a simple closed polygon $\mathfrak{p}^*$. Then, first of all, the construction used in §2.5 shows that we have a harmonic surface

$$\mathfrak{H}_0: x = x_0(u, v),\ y = y_0(u, v),\ z = z_0(u, v),\ u^2 + v^2 \leqq 1,$$

bounded by $\mathfrak{p}^*$ and having a finite area. Starting with *any* such surface $\mathfrak{H}_0$, the iterative process yields a sequence

$$\mathfrak{H}_n: x = x_n(u, v),\ y = y_n(u, v),\ z = z_n(u, v),\ u^2 + v^2 \leqq 1, \tag{40}$$

and $\mathfrak{H}_n$ solves the problem $P(\mathfrak{p}^*; A, B, C, A^*, B^*, C^*; \epsilon_n)$, where $\epsilon_1, \epsilon_2, \cdots, \epsilon_n, \cdots$ are positive numbers such that $\epsilon_n \rightarrow 0$. On account of the selection theorem, the sequence (40) contains a uniformly convergent subsequence, and the limit surface solves the problem of Plateau for $\mathfrak{p}^*$. It can easily be

† This result was first obtained by the author. See Radó, *The problem of the least area and the problem of Plateau*, Mathematische Zeitschrift, vol. 32 (1930), pp. 763–796. Subsequent proofs have been given by J. Douglas, *Solution of the problem of Plateau*, these Transactions, vol. 33 (1931), pp. 263–321, and, quite recently, by E. J. McShane, *Parametrizations of saddle surfaces, with application to the problem of Plateau*, in the current volume of these Transactions, pp. 716–733. All the proofs of this result known at the present time depend on the passage from S to $\mathfrak{H}$ (see the Intoduction) first used in the author's paper *On Plateau's problem*, Annals of Mathematics, (2), vol. 31 (1930), pp. 457–469.

shown that if we choose the initial harmonic surface $\mathfrak{H}_0$ in all possible ways, then we obtain in this way *all* the solutions of the problem of Plateau. Indeed let

$$S: x = x(u, v),\ y = y(u, v),\ z = z(u, v),\ u^2 + v^2 \leqq 1,$$

be any solution of the problem as stated in §1.14. Then $x(u, v)$, $y(u, v)$, $z(u, v)$ are harmonic in $u^2+v^2<1$, and thus we can use S as the initial harmonic surface of the iterative process, *provided* the area of S is finite. But this is indeed so, on account of a theorem of Carleman. According to Carleman, every minimal surface, of the topological type of the circular disc, satisfies the isoperimetric inequality

$$\mathfrak{A} \leqq \frac{1}{4\pi} L^2$$

where $\mathfrak{A}$ is the area of the surface and L is the length of its perimeter.† In our case, the perimeter is a polygon, and thus L and consequently $\mathfrak{A}$ is finite. Thus S can be used as the initial harmonic surface $\mathfrak{H}_0$ of the iterative process, and it is then obvious, on account of condition (b) of the problem of Plateau (see §1.14), that all the harmonic surfaces $\mathfrak{H}_n$ coincide with S.

4.4. Thus the fact that the iterative process yields all the solutions of the problem of Plateau appears as trivial. On the other hand, I feel that this fact constitutes the specific advantage of the method. My own previous work, as well as the work of Douglas and that of McShane, yielded a solution with a minimum area, and therefore certainly not the general solution.

The iterative process might be considered therefore as a contribution to *the problem of determining the totality of the solutions of the problem of Plateau.* If the given curve has a simply covered convex curve as its parallel or central projection upon some plane, then the solution of the problem is unique.‡ As far as I know, the exact number of the solutions has not yet been determined in any other case.

† See, for a simplified proof and literature, E. F. Beckenbach, *The area and boundary of minimal surfaces*, Annals of Mathematics, vol. 33 (1932), pp. 658–664. Further developments on the isoperimetric inequality are contained in a joint paper by E. F. Beckenbach and the present author, *Subharmonic functions and surfaces of negative curvature*, these Transactions, vol. 35 (1933), pp. 662–674.

‡ See Radó, Acta Szeged, vol. 6 (1932), pp. 1-20.

Ohio State University
Columbus, Ohio

EFFECTS OF LINEAR TRANSFORMATIONS ON THE DIVERGENCE OF BOUNDED SEQUENCES AND FUNCTIONS*

BY

JOSEPH LEV

1. **Introduction.** The transformation

$$y_n = \sum_{i=1}^{\infty} K_{n,i} x_i,$$

where $\{x_i\}$ is a sequence of complex elements and the $K_{n,i}$ are complex numbers, has been widely studied, and the conditions which must be fulfilled by the $K_{n,i}$ in order that the property of convergence of the sequence may remain invariant were given by Schur [1].† In recent studies by Hurwitz [2, 3] and Knopp [4] modes of measuring the divergence of bounded sequences were given, and the conditions on the $K_{n,i}$ were found under which the divergence of the sequence $\{y_n\}$ is no greater than that of $\{x_n\}$.

In this paper the effects of the transformations will be investigated with fewer restrictions on the $K_{n,i}$ than those imposed by earlier writers. The problem will be approached by means of the new concept of the *limit circle* defined as follows:

The limit circle of a bounded sequence of complex elements is the (unique) circle of least radius which contains within or on its boundary the limit points of the sequence.

The limit circle of a bounded function $F(y)$ of the complex variable y as $y \to \xi$ (finite or infinite) is analogously defined in terms of the limit points of $F(y)$ as $y \to \xi$; this concept will be used in the study of transformations of sequences and functions into functions.

2. **Sequence to function transformations.** Instead of the transformation mentioned in the introduction we shall study the following more general transformation S. Let T be a set of points in the complex plane having a limit point t_0 (finite or infinite) not belonging to T. We shall speak of a point t in T as being sufficiently advanced if for some $\delta > 0$, $|t - t_0| < \delta$ when t_0 is finite, or $|1/t| < \delta$ when t_0 is infinite. Then let $K_i(t)$ be a set of complex numbers defined for $i = 1, 2, \cdots$, and each t in T, and such that

* Presented to the Society, December 27, 1932; received by the editors November 15, 1932, and after revision, May 17, 1933.

† Here and below numbers in square brackets refer to the bibliography at the end of the paper.

$$S: \qquad g(t) = \sum_{i=1}^{\infty} K_i(t)x_i$$

is defined for each t in T. We shall refer to the limit points of $g(t)$ as $t \to t_0$ simply as the limit points of $g(t)$.

We shall now prove

THEOREM 2.1. *Let $\{x_n\}$ be a bounded sequence of complex elements. If the $K_i(t)$ satisfy the conditions*

$$\lim_{t \to t_0} K_i(t) = k_i, \text{ for each } i, \tag{2.11}$$

$$\sum_{i=1}^{\infty} | K_i(t) | < M, \tag{2.12}$$

for all sufficiently advanced t, M a constant, then the quantities α the center and D the radius of the limit circle of the function $\sum_{i=1}^{\infty} K_i(t)$,

$$A = \alpha - \sum_{i=1}^{\infty} k_i, \quad B = \sum_{i=1}^{\infty} k_i x_i, \quad C = \limsup_{t \to t_0} \sum_{i=1}^{\infty} | K_i(t) - k_i|$$

exist, and the limit points of $g(t)$ lie in the circle of center $H = Ah + B$, and radius $R = Cr + D|h|$, where h is the center and r the radius of the limit circle of $\{x_n\}$.

The existence of α, A, B, C, D is easy to establish and the details will not be given here. For the remainder of the proof write the inequality

$$\begin{aligned} \left| \sum_{i=1}^{\infty} K_i(t)x_i - \sum_{i=1}^{\infty} k_i x_i - h\left(\alpha - \sum_{i=1}^{\infty} k_i\right)\right| &\leqq \sum_{i=1}^{\infty} | K_i(t) - k_i | \cdot | x_i - h | \\ &+ \sum_{i=p+1}^{\infty} |K_i(t) - K_i| \cdot | x_i - h| + | h| \cdot \left| \sum_{i=1}^{\infty} K_i(t) - \alpha\right|. \end{aligned}$$

Choose $\epsilon > 0$, and p so great that for all $i > p |x_i - h| < r + \epsilon$. Then

$$\begin{aligned} &\limsup_{t \to t_0} \left| g(t) - \sum_{i=1}^{\infty} k_i x_i - h\left(\alpha - \sum_{i=1}^{\infty} k_i\right)\right| \\ &\leqq o(1) + (r + \epsilon) \cdot \limsup_{t \to t_0} \sum_{i=1}^{\infty} | K_i(t) - k_i | + | h | \cdot \limsup_{t \to t_0} \left| \sum_{i=1}^{\infty} K_i(t) - \alpha \right|, \end{aligned}$$

and since the inequality holds for all $\epsilon > 0$, the theorem follows.

The remaining theorems of this section will be seen to be in part consequences of Theorem 2.1. Notations already introduced will be freely used,

and $\{x_n\}$ will be taken bounded throughout the discussion. In particular h and r will in each case by taken to depend on $\{x_n\}$.

Theorem 2.1 easily yields the sufficiency of the following theorem of Schur [1].

THEOREM 2.2. *In order that S may be such that* $\lim_{t\to t_0} g(t)$ *exists whenever* $h=r=0$, *it is necessary and sufficient that the* $K_i(t)$ *satisfy* (2.11), *and* (2.12).

The theorem just stated can be generalized to

THEOREM 2.3. *Let N be a real non-negative constant. In order that the limit points of* $g(t)$ *shall lie in a circle of radius* $N|h|$, *whenever* $r=0$, *it is necessary and sufficient that the* $K_i(t)$ *satisfy the conditions* (2.11), (2.12), *and* $D\leqq N$.

The sufficiency follows from Theorem 2.1, and the necessity of the first two conditions from Theorem 2.2. For the necessity of the condition $D\leqq N$ we need only consider the special case $x_n=1$ $(n=1, 2, \cdots)$.

To supplement Theorem 2.3 we can give the following theorem which takes into account the position of the limit points of $g(t)$.

THEOREM 2.4. *Let N be a real non-negative constant. In order that S may be such that the limit points of* $g(t)$ *shall lie in a circle of center h and radius* $N|h|$, *whenever* $r=0$, *it is necessary and sufficient that the* $K_i(t)$ *satisfy the conditions* (2.11), (2.12), $D\leqq N$, $k_i=0$, *for all i, and* $\alpha=1$.

The proof readily follows from a consideration of Theorems 2.1 and 2.3, and, for the necessity of the two last conditions, the sequences $x_i=0$, $i\neq j$, $x_j=1$, and the sequence $x_i=1$ $(i, j=1, 2, \cdots)$.

Obviously the special case $N=0$ yields the well known conditions for regularity, namely the conditions under which $\lim_{t\to t_0} g(t)=\lim_{n\to\infty} x_n$.

We shall now give two theorems which are concerned with divergent sequences.

THEOREM 2.5. *Let Q be a real non-negative constant. In order that S may be such that the limit points of* $g(t)$ *shall lie in a circle of radius Qr whenever* $h=0$, *it is necessary and sufficient that the* $K_i(t)$ *satisfy the conditions* (2.11), (2.12), *and* $C\leqq Q$.

In the proof we encounter difficulty only in connection with establishing necessity of the condition $C\leqq Q$. We shall assume that (2.11) and (2.12) hold and show that the remaining condition also holds.

Suppose on the contrary $C>Q$. Then for some $\lambda>0$ we have

$$\sum_{i=1}^{\infty} | K_i(t) - k_i | > Q + 5\lambda$$

repeatedly as t approaches t_0. There exists, therefore, a sequence $\{t_p\}$ lying entirely in the range for which (2.12) is satisfied, and such that $\lim_{p\to\infty} t_p = t_0$; and an increasing sequence of integers $\{n_p\}$ for which the following inequalities hold:

$$\sum_{i=1}^{n_{p-1}+1} | K_i(t_p) - k_i | < \lambda, \quad \sum_{i=1}^{\infty} | K_i(t_p) - k_i | > Q + 5\lambda,$$

$$\sum_{i=n_p+1}^{\infty} | K_i(t_p) - k_i | < \lambda, \quad \sum_{i=n_{p-1}+2}^{n_p} | K_i(t_p) - k_i | > Q + 3\lambda.$$

In the set $K_i(t_p)-k_i$, $n_{p-1}+2 \leqq i \leqq n_p$, there is surely one value $K_s(t_p)-k_s$ which is not zero.

We now define a sequence having a limit circle of center zero and radius one*

$$x_i = (-1)^{p-1} \operatorname{sgn} [K_i(t_p) - k_i], \quad n_{p-1} + 2 \leqq i \leqq n_p,$$
$$x_{n_p+1} = (-1)^p \operatorname{sgn} [K_s(t_p) - k_s].$$

We shall establish the desired contradiction if we show that the limit circle of

$$g(t) = \sum_{i=1}^{\infty} K_i(t) x_i$$

has a radius greater than Q. Write

$$g(t_p) - \sum_{i=1}^{\infty} k_i x_i = \left(\sum_{i=1}^{n_{p-1}+1} + \sum_{i=n_{p-1}+2}^{n_p} + \sum_{i=n_p+1}^{\infty} \right) [K_i(t_p) - k_i] x_i.$$

The first and third terms on the right are each less than λ in absolute value, and the real middle term is greater than $Q+3\lambda$ for p odd and less than $-(Q+3\lambda)$ for p even. Hence, writing $R(z)$ = real part of z,

$$R\left[g(t_p) - \sum_{i=1}^{\infty} k_i x_i \right] > Q + \lambda, \ p \text{ odd},$$
$$< -(Q + \lambda), \ p \text{ even},$$

and $g(t)$ has a limit circle of radius greater than Q, which completes the proof.

The conditions (2.11), (2.12), and $C \leqq Q$, remain necessary but not sufficient when Theorem 2.5 is written without the hypothesis $h=0$. We can, however, state necessary and sufficient conditions if we restrict ourselves to a consideration of conservative transformations, that is, those transformations

* We use the definition $\operatorname{sgn}(z) = |z|/z$, $z \neq 0$, and $\operatorname{sgn}(z) = 0$, $z = 0$.

for which $\lim_{t\to t_0} g(t)$ exists whenever $\lim_{n\to\infty} x_n$ exists. The conditions for conservatism are (2.11), (2.12), and $D=0$. Clearly in the conservative case, the condition $C\leqq Q$ is necessary and sufficient for the limit points of $g(t)$ to lie in a circle of radius Qr. We can also take into account the position of the limit points and state

THEOREM 2.6. *Let Q be a constant, $Q\geqq 1$. In order that the conservative transformation S may be such that the limit points of $g(t)$ shall lie in a circle of radius Qr and center h, whenever $\{x_n\}$ is bounded, it is necessary and sufficient that S be regular, and that $C\leqq Q$.*

An example due to W. A. Hurwitz yields an interesting comparison between the work of this paper and that of Hurwitz and Knopp. Apply to the sequence $x_n=\omega^{2n}$, $\omega^3=1$, the transformation defined by $K_{n,i}=(-1)^n\omega^i/3$ $(i=n,n+1,n+2)$, and $K_{n,i}=0$, otherwise. The resulting sequence, $g_n=(-1)^n$, has its limit points within the limit circle of the original sequence as is to be expected from our theory but the oscillation of $\{g_n\}$ is greater than that of $\{x_n\}$ and one of its limit points lies outside the limit core of $\{x_n\}$.

3. **Function to function transformations. I.** In the following let $f(x)$ be a complex function of the real variable x defined and integrable Lebesgue in each interval $a\leqq x\leqq x_1<\xi$, where x_1 is arbitrary and ξ is finite or infinite.

We shall call the following the transformation S_1. Choose a point set T as in the definition of S, and a function $K_1(t, x)$ defined for each t in T, and each x, $a\leqq x<\xi$, integrable Lebesgue in each interval $a\leqq x\leqq x_1<\xi$, for each t, such that

$$S_1: \qquad g_1(t) = \int_a^\xi K_1(t, s)f(s)ds$$

exists for each t in T.

We shall now give without proof a theorem analogous to Theorem 2.1.

THEOREM 3.1. *Let $f(x)$ be bounded $a\leqq x<\xi$. If S_1 is such that $K_1(t, x)$ satisfies the conditions*

$$\lim_{u,t\to t_0} \int_a^{x_1} | K_1(t, s) - K_1(u, s) | ds = 0, \quad a \leqq x_1 < \xi, \tag{3.11}$$

$$\int_a^\xi | K_1(t, s) | ds < M, \tag{3.12}$$

for all sufficiently advanced t, M a constant, then the quantities α_1 the center and D_1 the radius of the limit circle of the function $\int_a^\xi K_1(t, s)ds$,

$$A_1 = \alpha_1 - \lim_{x\to\xi} \lim_{t\to t_0} \int_a^x K_1(t, s)ds,$$

$$B_1 = \lim_{x\to\xi} \lim_{t\to t_0} \int_a^x K_1(t, s)f(s)ds,$$

$$C_1 = \limsup_{t\to t_0} \lim_{x\to\xi} \lim_{u\to t_0} \int_a^x | K_1(t, s) - K_1(u, s) | ds$$

exist, and the limit points of $g_1(t)$ *lie in a circle of center* $H_1 = A_1h + B_1$, *and radius* $R_1 = C_1r + D_1|h|$, *where* h *is the center and* r *the radius of the limit circle of* $f(x)$.

The sufficiency of theorems analogous to those in §2 can easily be established, but for a complete theory analogous to that in §2 we need the transformations in the next section.

4. Function to function transformations. II. We shall call the following transformation S_2. Choose a function $K_2(t, x)$ which has all the properties of $K_1(t, x)$ and the additional property that $K_2(t, x)$ is continuous in x, uniformly for all sufficiently advanced t, and all x, $a \leqq x \leqq q$, where q is an arbitrary constant less than ξ. The transformation is then given by

$$S_2: \qquad g_2(t) = \int_a^\xi K_2(t, s)f(s)ds.$$

We can establish

THEOREM 4.1. *Let* $f(x)$ *be bounded,* $a \leqq x < \xi$. *If* S_2 *is such that* $K_2(t, x)$ *satisfies the conditions*

$$\lim_{t\to t_0} K_2(t, x) = k(x),\ a \leqq x < \xi, \tag{4.11}$$

$$\int_a^\xi | K_2(t, s) | ds < M, \tag{4.12}$$

for all sufficiently advanced t, M *a constant, then the quantities* α_2 *the center and* D_2 *the radius of the limit circle of the function* $\int_a^\xi K_2(t, s)ds$,

$$A_2 = \alpha_2 - \int_a^\xi k(s)ds, \quad B_2 = \int_a^\xi k(s)f(s)ds,$$

$$C_2 = \limsup_{t\to t_0} \int_a^\xi | K_2(t, s) - k(s) | ds$$

exist, and the limit points of $g_2(t)$ *lie in a circle of center* $H_2 = A_2h + B_2$, *and radius* $R_2 = C_2r + D_2|h|$, *where* h *is the center and* r *the radius of the limit circle of* $f(x)$.

The proof of this theorem may be made to depend upon that of Theorem 3.1 by showing that $k(x)$ is continuous $a \leqq x < \xi$, that $K_2(t, x)$ approaches $k(x)$ uniformly over $a \leqq x \leqq q$, for arbitrary q less than ξ, and that $k(x)$ is integrable over $a \leqq x \leqq \xi$. The details will not be given here.

We can now state

THEOREM 4.2. *In order that S_2 may be such that $\lim_{t \to t_0} g_2(t)$ exists whenever $h = r = 0$, it is necessary and sufficient that $K_2(t, x)$ satisfy the conditions* (4.11) *and* (4.12).

The sufficiency follows from Theorem 4.1. The necessity can be established by using the methods of Silverman [5], and Schur [1]; and a consideration of the fact that if $f(x)$ is measurable $a \leqq x \leqq b$, then sgn $f(x)$ is also measurable in this interval.

The remaining analogues of the theorems in §2 can easily be stated and proved by methods suggested in that section, and will not be given here.

5. **Bounds of the sets of limit points.** It is easy to see that in parts of our discussion we can replace the limit circle by some other circle which contains the limit points of the sequence or function. In particular we can replace it by a circle with center at the origin and radius equal to the maximum of the distances from the origin to the limit points. This radius which is a bound for the set of limit points may be written in the case of sequences as $\limsup_{n \to \infty} |x_n|$. We can state

THEOREM 5.1. *Let Q be a real non-negative constant. In order that*

$$\limsup_{t \to t_0} |g(t)| \leqq Q \limsup_{n \to \infty} |x_n|$$

whenever $\{x_n\}$ is bounded, it is necessary and sufficient that the $K_i(t)$ satisfy the conditions (2.11), (2.12), *$C \leqq Q$, and $k_i = 0$, for all i.*

For the proof of necessity consider Theorem 2.5 and, for the last condition, the sequence $x_i = 0$ $(i = 1, 2, \cdots)$; and the sequences $x_j = 1$, $x_i = 0$, $i \neq j$ $(i, j = 1, 2, \cdots)$.

6. **Application to series.** We shall generalize some results due to Schur [1] and Kojima [6].

Let the series $w_0 + w_1 + w_2 + \cdots$, with partial sums W_n, be the Cauchy product of the two series $u_0 + u_1 + u_2 + \cdots$, and $v_0 + v_1 + v_2 + \cdots$, with partial sums U_n and V_n respectively. We can write

$$W_n = u_n V_0 + \cdots + u_0 V_n,$$

which is a linear transformation on the sequence $\{V_n\}$. If we suppose that $\sum |u_n|$ converges and that $\{V_n\}$ is bounded with limit circle of center h and

radius r we can apply Theorem 2.1 to show that the limit points of $\{W_n\}$ lie in a circle of center $h\sum u_n$ and radius $r\sum|u_n|$.

Now write $U_n^p=\sum_{i=0}^n A_{n-i}^p u_i$, where $A_n^p=\Gamma(p+n+1)/[\Gamma(n+1)\Gamma(p+1)]$ and $p\geqq 0$. If the Cesàro transform of order p, $C_n^p(u)=U_n^p/A_n^p$, is bounded, the series $\sum u_n$ is said to be bounded (C, p). Writing similar expressions for the series $\sum v_n$ and $\sum w_n$ we get $W_n^{p+q+1}=\sum_{i=0}^n V_{n-i}^q U_i^p$, $q\geqq 0$, which may be written

$$C_n^{p+q+1}(w)=\frac{1}{A_n^{p+q+1}}\sum_{i=0}^n A_i^p V_{n-i}^q C_i^p(u).$$

If we regard this expression as a linear transformation on the $C_i^p(u)$ we get $K_{n,i}=A_i^p V_{n-i}^q/A_n^{p+q+1}$ $(i=0, 1, 2, \cdots, n)$, $K_{n,i}=0$ $(i>n)$. Then supposing that $\sum v_n$ is bounded (C, q), we have $|K_{n,i}|=A_i^p A_{n-i}^q|C_{n-i}(v)|/A_u^{p+q+1}<M/n^{p+1}$, M a constant for all n, so that $\lim_{n\to\infty} K_{n,i}=0$. Furthermore $\sum_{i=0}^n|K_{n,i}|<N$, N a constant for all n, and $\sum_{i=0}^n K_{n,i}=C_n^{p+q+1}(v)$. Hence if we call the center and radius of $\{C_n^p(u)\}$, and $\{C_n^{p+q+1}(v)\}$, h_u, r_u, and h_v, r_v, respectively, we have

THEOREM 6.1. *If $\sum u_n$ is bounded (C, p) and $\sum v_n$ is bounded (C, q), $p, q\geqq 0$, then the sequence $\{C_n^{p+q+1}(w)\}$ has its limit points in a circle of center $h_v h_u$, and radius $r_u\cdot\limsup_{n\to\infty}\sum_{i=0}^n A_i^p|V_{n-i}^q|/A_n^{p+q+1}+r_v|h_u|$.*

If we consider the two series $\sum u_n$ with partial sums s_n, and $\sum c_n u_n$ with partial sums t_n we get

$$t_n = s_0(c_0-c_1)+\cdots+s_{n-1}(c_{n-1}-c_n)+s_n c_n.$$

On the basis of the assumptions that $\{s_n\}$ is bounded with limit circle of center h and radius r, and that $\sum_{n=0}^\infty|c_n-c_{n+1}|$ converges, we can show by means of Theorem 2.1 that the limit points of $\{t_n\}$ lie in a circle of center $h\cdot\lim_{n\to\infty} c_n+\sum_{n=0}^\infty(c_n-c_{n+1})s_n$, and radius $r\cdot\lim_{n\to\infty}|c_n|$.

Generalizations of the last result to the case when $\sum u_n$ is bounded (C, p) for some $p>0$ can easily be arrived at on the basis of the work of Schur [1] and Kojima [6].

BIBLIOGRAPHY

1. J. Schur, *Über lineare Transformationen in der Theorie der unendlichen Reihen*, Journal für die reine und angewandte Mathematik, vol. 151 (1920), pp. 79–111.

2. W. A. Hurwitz, *Some properties of methods of evaluation of divergent sequences*, Proceedings of the London Mathematical Society, (2), vol. 26 (1927), pp. 231–248.

3. W. A. Hurwitz, *The oscillation of a sequence*, American Journal of Mathematics, vol. 52 (1930), pp. 611–616.

4. K. Knopp, *Zur Theorie der Limitierungsverfahren*, Mathematische Zeitschrift, vol. 31 (1930), pp. 97–127, and 276–305.

5. L. L. Silverman, *On the notion of summability for the limit of a function of a real variable*, these Transactions, vol. 17 (1916), pp. 284–294.

6. T. Kojima, *On generalized Toeplitz's theorem on limits*, Tôhoku Mathematical Journal, vol. 12 (1917), pp. 291–326.

Cornell University,
Ithaca, N. Y.

GROUPS IN WHICH EVERY OPERATOR HAS AT MOST A PRIME NUMBER OF CONJUGATES*

BY
G. A. MILLER

Let G represent a non-abelian group such that each of its operators is either invariant or has p conjugates, p being a prime number which is the same for every operator of G, and let s_1 represent one of the operators of lowest order contained in G and non-invariant under G. The subgroup H_1 composed of all the operators of G which are commutative with s_1 is of index p under G, and includes s_1 as well as the central H_0 of G. It will be proved first that G involves only one Sylow subgroup of order p^m. If it could contain more than one subgroup of order p^m the powers of some operator of one of these would transform the cross-cut of two of them into itself and would also transform one of these two subgroups into itself and the other into exactly p distinct subgroups, since this operator may be so selected that its pth power is in this cross-cut but it itself is not contained therein.

To prove that this operator would have more than p conjugates under G it is only necessary to observe that it would appear in one and only one of a set of p subgroups of order p^m which would be conjugate under the powers of some operator contained in one of the given $p+1$ subgroups of this order. As these p subgroups would contain an operator which would transform the given operator into at least one additional conjugate, the following theorem has been established:

If a group contains more than one Sylow subgroup of order p^m then it contains a set of conjugate operators whose order is a power of p and whose number exceeds p, where p is any prime number.

By hypothesis G transforms the operators contained in each set of conjugates according to a transitive permutation group of degree p and it has just been proved that this transitive group cannot involve more than one subgroup of order p since such a subgroup is a Sylow subgroup therein. As G must be isomorphic with every such transitive group it follows directly that each of these transitive groups is cyclic. Hence G must be isomorphic with an abelian group of type $(1, 1, 1, \cdots)$ whose order is a power of p. That is, G is the direct product of a non-abelian group of order p^m and an abelian group whose order is prime to p. In what follows it may therefore be

* Presented to the Society, June 19, 1933; received by the editors March 13, 1933.

assumed for the sake of simplicity that the order of G is p^m. Since every subgroup of index p under a group of order p^m is invariant thereunder it follows that H_1 is an invariant subgroup of G.

An operator of G which is not found in H_1 transforms s_1 into s_0s_1 where s_0 is commutative with s_1 since H_1 is an invariant subgroup of G and all of its operators are commutative with s_1. Hence it results that s_0 is of order p. To prove that s_0 appears in the central of G it may be noted that it must be transformed into itself by each of the operators which do not appear in H_1 since every such operator must be commutative with a subgroup of index p under H_1 and s_0 must be in such a subgroup since it arises from a p automorphism. If t were a non-invariant operator of G which would not be transformed into itself multiplied by a power of s_0, then G would involve operators which would be commutative neither with s_1 nor with t. Such an operator would therefore be transformed into more than p conjugates under G. It has therefore been proved that *if every non-invariant operator of a group has a given prime number of conjugates under this group then the order of the commutator subgroup of this group is this prime number and its operators are invariant under the group.*

When H_1 is non-abelian it contains an invariant subgroup H_2 composed of all of its operators which are commutative with one, s_2, of its operators of lowest order. By continuing this process we finally arrive at an invariant abelian subgroup H_λ which involves $s_1, s_2, \cdots, s_\lambda$ as well as H_0. A set of independent generators of H_λ can be so selected that it includes $s_1, s_2, \cdots, s_\lambda$ since these were always chosen so as to be of the lowest possible order. As H_0 includes the pth power of every operator of G it includes the pth powers of $s_1, s_2, \cdots, s_\lambda$ and the central quotient group of G is abelian and of type $(1, 1, 1, \cdots)$. In particular, G/H_λ has these properties. It should be noted that H_0 is the direct product of the group generated by the pth powers of $s_1, s_2, \cdots, s_\lambda$ and some other group which may be the identity. The order of G/H_0 is $p^{2\lambda}$ while that of G/H_λ is p^λ.

To exhibit the fact that the preceding theorems relate to an extensive category of groups it may be noted that for every value of $m>2$ there are groups which belong to this category and that the number of these groups increases with m. In particular, the two non-abelian groups of order p^3 belong to this category and six of the non-abelian groups of order p^4 belong thereto. When $p=2$ it is known that there are nine non-abelian groups of this order and when $p>2$ their number is always ten. Hence more than one-half of the non-abelian groups of order p^4 come under the heading of the present article. For all of these groups $\lambda=1$ since H_0 cannot be the identity and the order of

G/H_0 is always $p^{2\lambda}$, as was noted above. Whenever $m>4$ then λ can obviously have more than one possible value.

When m is given and greater than 2 the possible values of λ depend upon the type of the abelian group which is selected for H_λ since λ may assume successively the values $1, 2, \cdots$ up to the number of the invariants of this abelian group if at least one of these invariants exceeds p. When all of these invariants are equal to p the value of λ may be any positive integer which does not exceed the number of these invariants diminished by unity in view of the following theorem:

If a non-abelian group G in which all the non-invariant operators have exactly a prime number p conjugates contains a non-invariant operator s_1 of order p, and if the subgroup composed of all of its operators which are commutative with s_1 is either abelian or contains a non-invariant operator s_2 of order p, etc., then the operators $s_1, s_2, \cdots, s_\lambda$ generate an abelian group of order p^λ which does not include the commutator subgroup of G.

To prove this theorem it is only necessary to note that this abelian subgroup contains no operator besides the identity which is invariant under G.

To construct all the groups of order p^m which satisfy the conditions imposed on G at the opening of this article we may first consider those in which $\lambda=1$, then those in which $\lambda=2$, etc., until we arrive at those in which λ has its largest possible value, viz., $(m-1)/2$ when m is odd and $(m-2)/2$ when m is even. For the central of such groups we may take successively every possible abelian group of order $p^{m-2\lambda}$, and two groups in which the centrals are distinct groups must themselves be distinct, so that we may avoid duplicates by classifying these groups of the same order according to their distinct centrals. When H_λ is cyclic, $\lambda=1$ since H_λ involves $s_1, s_2, \cdots, s_\lambda$ as independent generators. Moreover, G must then be the quaternion group since H_1 is of index p and G must then involve $p+1$ cyclic subgroups of this index. It is well known that the quaternion group is the only group of order p^m, $m>2$, which involves $p+1$ cyclic subgroups of index p.

When $\lambda=1$ general formulas for the totality of the non-abelian groups which come under the heading of the present article may be obtained as follows. Suppose first that all the invariants of H_0 are equal to a fixed number p^α. It is known that if any abelian group has a subgroup of prime index then it is always possible to find a set of reduced independent generators of this group which has the property that all the operators of this set except one appear in this subgroup.* Hence a set of reduced independent generators of

* G. A. Miller, Bulletin of the American Mathematical Society, vol. 23 (1916), p. 14. The term "reduced set of independent generators" was used with its present meaning with respect to abelian groups in these Transactions, vol. 16 (1915), p. 22.

each of the $p+1$ abelian subgroups of index p contained in G can be so selected that all except one of them appear in H_0. The additional independent generator of such a reduced set can therefore be so selected in the present case that its order is either p or $p^{\alpha+1}$. In what follows it will be assumed that such a selection of a set of the reduced independent generators of these subgroups has been made.

When H_0 is cyclic there are always two possible groups. In the special case when $p=2$ and $m=3$ these are the octic and the quaternion groups while in all the other cases the additional generators of two abelian subgroups of index p under G can be so selected that either both are of order p or one is of order p and the other is of order p^{m-1}. When H_0 has two equal invariants p^α there are four groups. In one of these the additional generators of two of the abelian subgroups of index p can be so chosen that both are of order p. In two others one of these generators is of order p while the other is of order $p^{\alpha+1}$ except when $p=2$ and $m=4$. In this special case there is only one such additional group while there are two groups in which all the additional generators are of order $p^{\alpha+1}$. In the other cases there is only one such group. When the number of the equal invariants of H_0 exceeds two there is one additional group in which all the additional independent generators are of order $p^{\alpha+1}$ since the commutator subgroup for such groups can then be chosen in two distinct ways. When $p=2$ and $m=4$ this commutator subgroup can be chosen in three different ways but there are then only two distinct groups under the other cases.

For the sake of simplifying the consideration of the general case when $\lambda=1$ we let k_1 represent the number of the different values of the invariants of H_0, k_2 the number of the sets composed separately of all the equal invariants whenever such a set involves at least two such invariants, k_3 the number of such sets such that each set involves at least three equal invariants. The number of the distinct groups of order p^m which contain this H_0 and in which the additional invariant of each of two abelian subgroups of index p is equal to p is then k_1, since H_0 contains k_1 sets of subgroups of order p such that each set is composed of all of its subgroups of this order which are conjugate under the holomorph of H_0.

If one of the $p+1$ abelian subgroups of index p under G has an invariant which is equal to p but is not included among the invariants of H_0, while another has a larger invariant having these properties, there are $k_1^2+k_2$ groups of order p^m since the commutator subgroup may be taken from any one of the k_1 sets of conjugate subgroups of order p under the holomorph of H_0 and the second independent generator which does not appear among the independent generators of H_0 may have its pth power in any one of the k_1

sets of conjugate operators such that each set is composed of all the operators which can be separately used as an independent generator of H_0. In the special case when at least one of the invariants of H_0 is 2, one of the groups of this case is included among the k_1 groups defined in the preceding paragraph. Hence there are then only $k_1^2+k_2-1$ additional groups.

Finally, when none of the additional invariants is equal to p the pth powers of the independent generators of the $p+1$ abelian subgroups of index p under G which are not also independent generators of H_0 are equal to distinct independent generators of H_0 except possibly when $p=2$ and these additional invariants are equal to 4. In the latter case these operators of order 4 may generate the quaternion group and then the commutator subgroup of G is the subgroup of order 2 contained in this quaternion group. In all other cases two equal additional independent generators can be selected in k_2 essentially different ways while two such unequal generators can be selected in $k_1(k_1-1)/2$ essentially different ways.

In the former case the number of the distinct groups when none of these invariants is equal to 2 is $k_1k_2+k_3$ since in k_3 cases the subgroups of order p in H_0 which are conjugate under its holomorph are not conjugate in the holomorph of G. In the latter case the number of these groups is $k_1^2(k_1-1)/2$ $+k_2(k_1-1)$. Hence the total number of the distinct G's when $\lambda=1$ and H_0 does not involve an invariant which is equal to 2 is

$$k_1 + k_1^2(k_1 + 1)/2 + 2k_1k_2 + k_3.$$

This formula gives also the correct number of groups when H_0 involves at least one invariant which is equal to 2. It was noted above that in this case the number of groups in which only one additional invariant of the $p+1$ abelian subgroups of index p under G is p is one less than in the other cases, but the number of the groups in which all the additional invariants are equal to p^2 is then one more than in the other cases in view of the existence of the quaternion group. When H_0 involves more than one invariant which is equal to 2 and the commutator of order 2 is the square of one of the additional independent generators of order 4, and has a different square, then the operators of order 2 in the group generated by these operators of order 2 are not conjugate under its holomorph while they are thus conjugate in the other cases. This however does not affect the number of the possible distinct groups in this case.

If a group of order p^m contains more than one abelian subgroup of index p then the cross-cut of two such subgroups is its central and its commutator subgroup is of order p. Hence such a group belongs to the category of groups defined by the heading of the present article and the value of λ in this case

is unity. The formula given in the preceding paragraph therefore gives also the number of these distinct groups whenever we use successively for H_0 the different possible abelian groups of order p^{m-2}. Since every group of order p^4 contains at least one abelian subgroup of order p^3, it results that the only non-abelian groups of order p^4 which are not enumerated by the given formula are those which contain only one abelian subgroup of index p. There are three such groups when $p=2$, but whenever $p>2$ there are four such groups.

University of Illinois,
Urbana, Ill.

POLYNOMIAL DIOPHANTINE SYSTEMS*

BY

E. T. BELL

1. Formal definitions of polynomial diophantine systems, as understood in this paper, are given in §5, after the necessary algebraic preliminaries in §2 and certain arithmetical considerations in §4. The algebraic structure of the systems is stated in §3. Roughly, the systems are of the following type.

As always henceforth, let n denote an arbitrary constant (finite) integer >1. Let $a, b, \cdots, c$ be mn (m finite, $\geqq 1$) integers >0, and let $P_\alpha, Q_\beta, \cdots, R_\gamma$ ($\alpha=1, \cdots, a$; $\beta=1, \cdots, b$; $\cdots$; $\gamma=1, \cdots, c$) be polynomials in sn independent variables $x_1, \cdots, x_{sn}$ (s finite, $\geqq 1$) with integer (not necessarily rational integer) coefficients. Let the system Σ,

$$\Sigma: P_1 = \cdots = P_a, Q_1 = \cdots = Q_b, \cdots, R_1 = \cdots = R_c,$$

be consistent and indeterminate.

The systems Σ considered have an infinity of integer solutions, all of which can be given explicitly by expressing $x_1, \cdots, x_{sn}$ as polynomials in parameters ranging over all integers, with integer coefficients, and for this complete solution only an application of the fundamental theorem of arithmetic (unique prime decomposition) is necessary. Homogeneous and inhomogeneous systems Σ are treated by the same analysis, and the degrees of the polynomials are unrestricted. The simple algorithm for obtaining the complete solution in integers is indicated in §7, with examples.†

The remarkable features are the complete, explicit, solvability and the intimate connection with the fundamental theorem. In these respects systems Σ are an immediate generalization in one direction of the Pythagorean equation $x^2+y^2=z^2$ and its complete solution in rational integers. Naturally, the polynomials composing a system Σ can not be given arbitrarily; applicability of the fundamental theorem imposes necessary restrictions. We proceed to the algebra sufficient for the construction of general systems Σ.

2. Let R be an abstract commutative ring in which the identities with respect to addition, multiplication are z, u respectively, and in which the sum, product of any elements a, b in R are written $a+b$, ab respectively.

---

* Presented to the Society, October 28, 1933; received by the editors May 26, 1933.

† In all of the examples constructed (only 3 are reproduced here) all of the polynomials in the systems are irreducible in the ring of their values, but I can not prove that all polynomials in the most general system constructed are irreducible.

Since R may contain nilfactors, $ax=bx$, $x\neq z$ do not necessarily imply $a=b$. In the special case when R is a domain of integrity, the cancellation law holds. Conversely, in the special case when cancellation holds, R is a domain of integrity. For if $p\neq z$, $q\neq z$, $pq=z$, then $pq=pz$, and hence $q=z$, a contradiction. Unless so stated it is not assumed that R (or any other commutative ring) is a domain of integrity.

The elements of R will be called *integers*; to avoid confusion, 0, ± 1, ± 2, $\cdots$ will always be characterized as rational integers.

Integers other than z, u will frequently be indicated by a multiple index notation, $x(i, j)$, $y(k, t)$, $x(i_1, \cdots, i_s; j)$, the i, j, k, t being the indices. In a symbol with precisely 2 indices, say $x(i, j)$, the first index (i) ranges over all rational integers >0, the second (j) ranges over only $1, \cdots, n$.

A symbol with precisely 1 index, say $x(i)$, denotes a *vector* (one-rowed matrix) of n integers,

$$x(i) = (x(i, 1), \cdots, x(i, n));$$

the jth *coordinate* of $x(i)$ is $x(i, j)$. Vectors being matrices, vector equality, $x(i)=y(k)$, is matrix equality, $x(i, j)=y(k, j)$ $(j=1, \cdots, n)$.

It is postulated that there exists in R a set ϕ of integers $\phi(i, j, k)$ $(i, j, k=1, \cdots, n)$ such that

$$\phi(1, s, k) = \delta_{sk} \qquad (s, k = 1, \cdots, n) \tag{2.1}$$

$$[\delta_{ss} = u,\ \delta_{sk} = z,\ s \neq k];$$

$$\phi(i, j, k) = \phi(j, i, k) \qquad (i, j, k = 1, \cdots, n); \tag{2.2}$$

$$\sum_{j=1}^{n}\phi(p, q, j)\phi(j, r, t) = \sum_{j=1}^{n}\phi(p, r, j)\phi(j, q, t) \quad (p, q, r, t = 1, \cdots, n). \tag{2.3}$$

Let c_{pj} $(p, j=1, \cdots, n)$ be integers, such that $c_{1j}=\delta_{1j}$ $(j=1, \cdots, n)$ and the determinant $|c_{pj}|$ (p row, j column) of the matrix $\|c_{pj}\|$ has the value u. Let c'_{pj} denote the cofactor of c_{pj} in $|c_{pj}|$. Then $c'_{pj}=u$, and

$$\sum_{j=1}^{n} c_{pj}c'_{rj} = \delta_{pr} = \sum_{j=1}^{n} c_{jp}c'_{jr} \qquad (p, r = 1, \cdots, n).$$

Define the set ϕ' of integers $\phi'(r, s, t)$ $(r, s, t=1, \cdots, n)$ by

$$\phi'(r, s, t) \equiv \sum_{j,k,h=1}^{n} c_{rj}c_{sk}c'_{th}\phi(j, k, h).$$

Then it is easily seen (by manipulation of dummy suffixes as in tensor algebra) that

$$\phi(r, s, t) = \sum_{j,k,h=1}^{n} c'_{jr}c'_{ks}c_{ht}\phi'(j, k, h).$$

Similarly, and by using (2.1)–(2.3), we see that the $\phi'(r, s, t)$ satisfy the same relations: the symbol ϕ in (2.1)–(2.3) can be replaced by ϕ'. We shall say that the sets ϕ, ϕ' are *equivalent*, $\phi\sim\phi'$. The relation of equivalence is reflexive ($\phi\sim\phi$), symmetric (if $\phi\sim\phi'$ then $\phi'\sim\phi$), and transitive (if $\phi\sim\phi'$ and $\phi'\sim\phi''$, then $\phi\sim\phi''$). The consistency of (2.1)–(2.3) need not be discussed, as instances of ϕ will be evident when diophantine systems are constructed.

The jth coordinate, denoted by $x(i_a, i_b; j)$, in the *product* $x(i_a)x(i_b)$ of any vectors $x(i_a)$, $x(i_b)$, *to the base* ϕ, is defined by

$$(2.4)\qquad x(i_a, i_b; j) = \sum_{j_a, j_b=1}^{n} \phi(j_a, j_b; j)x(i_a, j_a)x(i_b, j_b) \quad (j = 1, \cdots, n).$$

In a given context the same base ϕ is presupposed. All equations persist if ϕ be replaced by ϕ', where $\phi\sim\phi'$.

The notation $u(i)(=u(k), \cdots)$ is reserved for the vector defined by $u(i, j) = \delta_{ij}(j=1, \cdots, n)$. Hence, by (2.1), (2.4) we have

$$(2.5)\qquad u(i)x(k) = x(k)$$

for all $x(k)$. If possible, let $v(i)x(k) = x(k)$, $v(i) \neq u(i)$, all $x(k)$. Choosing $x(k) = u(k)$, and referring to (2.5), we have the contradiction $v(i) = u(i)$. Hence $u(i)$ is the unique identity of vector multiplication.

The notation $z(i)(=z(k), \cdots)$ is reserved for the vector defined by $z(i, j) = z(j=1, \cdots, n)$; $z(i)x(k) = z(i)$ for all $x(k)$.

The *sum* $x(i_a) + x(i_b)$ is the vector whose jth coordinate is the sum (in R) of the jth coordinates of $x(i_a)$, $x(i_b)$ $(j=1, \cdots, n)$. Since z is the unique identity of addition in R, $z(i)$ is the unique identity of vector addition.

There can be no confusion between operations in R and the corresponding operations on vectors, since the notation for the operands indicates the species.

Now (2.2), (2.3) are necessary and sufficient conditions for commutativity and associativity of multiplication in any linear algebra with n basal units, and (2.1) is a necessary and sufficient condition for the existence of an identity with respect to multiplication in the algebra. Since any commutative, associative linear algebra with an identity of multiplication has a vector representation with vector multiplication and addition as above defined, it follows that *the set, VR, of all vectors is a commutative ring, in which the identities of multiplication, addition are $u(i)$, $z(i)$ respectively.*

The notation $\epsilon(i)$, $\epsilon(k)$, $\cdots$ will be reserved for units in VR, which are defined as follows, and which are not to be confused with the usual unit vectors of linear algebra. If $\epsilon(i_1)$ is in VR, and if $\epsilon(i_2)$ exists in VR such that

$$\epsilon(i_1)\epsilon(i_2) = u(i),$$

$\epsilon(i_1)$ is a *unit* in VR, and $\epsilon(i_2)$ is its *conjugate*. Hence the conjugate of a unit is a unit. Since $u(i)$ is its own conjugate, units exist. If possible, let $\epsilon(i)=z(i)$. Then $\epsilon(i)\epsilon'(i)=z(i)\epsilon'(i)$, where $\epsilon'(i)$ is the conjugate of $\epsilon(i)$. Hence the contradiction (in R) $u=z$. Thus no unit is equal to the zero in VR.

Suppose for a monent that VR is a domain of integrity. If possible, let

$$\epsilon(i_1)\epsilon(i_2) = u(i) = \epsilon(i_1)\epsilon(i_3),\ \epsilon(i_2) \neq \epsilon(i_3).$$

Then $\epsilon(i_1)[\epsilon(i_2)-\epsilon(i_3)]=z(i)$. With $\epsilon(i_1)\neq z(i)$ this gives the contradiction $\epsilon(i_2)=\epsilon(i_3)$. Hence, if VR is a domain of integrity, the conjugate of a given unit is unique.

We return to the general VR. In order that $\epsilon(i_1)$, $\epsilon(i_2)$ be conjugate units it is necessary and sufficient that

$$\sum_{r,s=1}^{n} \phi(r, s, j)\epsilon(i_1, r)\epsilon(i_2, s) = \delta_{1j} \qquad (j = 1, \cdots, n).$$

Let $a(i)$, $\alpha(i)$, $b(i)$, $\beta(i)$, $\xi(i)$ be such that

$$a(i) = \alpha(i)\xi(i),\ b(i) = \beta(i)\xi(i).$$

If $a(i)$, $b(i)$ are given, a solution of these equations is of the type

$$\alpha(i) = a(i)\epsilon(i),\ \beta(i) = b(i)\epsilon(i),\ \xi(i) = \epsilon'(i),$$

where $\epsilon(i)$, $\epsilon'(i)$ are arbitrary conjugate units. If this type exhausts the solutions, $a(i)$, $b(i)$ are said to be *coprime* (in VR). Necessary and sufficient conditions that $a(i)$, $b(i)$ be coprime are that the system

$$\begin{aligned} &\sum_{r,s=1}^{n} \phi(r, s, j)\alpha(i, r)\xi(i, s) = a(i, j), \\ (2.6)\qquad &\sum_{r,s=1}^{n} \phi(r, s, j)\beta(i, r)\xi(i, s) = b(i, j), \\ &\sum_{r,s=1}^{n} \phi(r, s, j)\xi'(i, r)\xi(i, s) = \delta_{1j} \qquad (j = 1, \cdots, n) \end{aligned}$$

be solvable in R for

$$\alpha(i, r),\ \beta(i, r),\ \xi(i, s),\ \xi'(i, r) \qquad (r, s = 1, \cdots, n).$$

3. Let $x(1), \cdots, x(s)$ be any vectors, and let $s>2$. The jth coordinate in the product $x(1) \cdots x(s)$ will be denoted by $x(1, \cdots, s; j)$. By mathematical induction from (2.4) we find for $x(1, \cdots, s; j)$ the following explicit polynomial expression in R:

$$\sum\phi(j_1, j_2, k_1)\phi(k_1, j_3, k_2)\cdots\phi(k_{s-3}, j_{s-1}, k_{s-2})\phi(k_{s-2}, j_s, j)$$
$$\times\, x(1, j_1)x(2, j_2)\cdots x(s, j_s)\quad (j_1, \cdots, j_s, k_1, \cdots, k_{s-2} = 1, \cdots, n).$$

When $x(i)=x(i_1)=\cdots=x(i_t)(t>1)$ the product $x(i_1)\cdots x(i_t)$ is written $x^t(i)$, and its jth coordinate $x(i^{(t)}; j)$. Hence, the case $t=2$, $s_1=s_2=1$, included, we have defined $x^{s_1}(i_1)\cdots x^{s_t}(i_t)$ and its jth coordinate $x(i_1^{(s_1)}, \cdots, i_t^{(s_t)}; j)$.

If the n coordinates of $x(i)$ are independent variables in R, $x(i)$ is called a *variable* (vector) in VR. The variables $x(i_1), \cdots, x(i_t)$ in VR are said to be *independent* if their nt coordinates are nt independent variables in R. Denote the power product $x^{s_1}(i_1)\cdots x^{s_t}(i_t)$, where $s_1, \cdots, s_t$ are rational integers >0, of t independent variables $x(i_1), \cdots, x(i_t)$ in VR by $X(t)$, with a similar notation for any product of positive integral powers of independent variables in VR. The r power products $X_1(t_1), \cdots, X_r(t_r)$ $(r>1)$ in VR are said to be *independent* if all the variables in VR composing these r products are independent in VR.

Denote the jth coordinate of $X_i(t_i)$ by $X_i(t_i; j)$, and let $X_1(t_1), \cdots, X_r(t_r)$ be independent. Then the equations

$$X_1(t_1) = \cdots = X_r(t_r) \tag{3.1}$$

in VR are equivalent to the simultaneous system

$$X_1(t_1; j) = \cdots = X_r(t_r; j)\qquad (j = 1, \cdots, n) \tag{3.2}$$

in R, as each of (3.1), (3.2) implies the other. With $a, b, \cdots, c, m$ as in §1, we pass to the general case. The power products in each row of

$$\begin{aligned} X_1(p_1) &= \cdots = X_a(p_a),\\ Y_1(q_1) &= \cdots = Y_b(q_b),\\ \cdots\ &\qquad \cdots\\ Z_1(r_1) &= \cdots = Z_c(q_c) \end{aligned} \tag{3.3}$$

are independent; in any pair of rows, at least one product in one of the rows and one product in the other are not independent; the system (3.3) does not separate into two or more systems with the two preceding characteristics in sets of independent variables in VR having no variable in common. The set

$$\begin{aligned} X_1(p_1; j) &= \cdots = X_a(p_a; j),\\ Y_1(q_1; j) &= \cdots = Y_b(q_b; j),\\ \cdots\ &\qquad \cdots\\ Z_1(r_1; j) &= \cdots = Z_c(r_c; j)\qquad (j = 1, \cdots, n) \end{aligned} \tag{3.4}$$

in R, equivalent to (3.3) in VR, will be called a *polynomial system.*

If the complete solution in integers of a polynomial system is obtainable in explicit form in terms of polynomials in integer parameters with integer coefficients, we shall say the system is *diophantine*.

It will be seen that a sufficient condition that a polynomial system in R be diophantine is that the fundamental theorem of arithmetic shall hold in VR. A generalization of (3.4) including arbitrary constant coefficients is noted in §7.

4. Unique decomposition is understood here in the strict sense, as in rational arithmetic or in the theory of ideals in an algebraic number field. For precision the postulates are stated. The notation in this section is independent of that in the rest of the paper.

Let Ω denote a set of at least two distinct elements $a, b, \cdots$, for which the postulates (4.1)–(4.7) hold.

(4.1) Equality is significant in Ω; $a=b$ or $a\neq b$; equality is symmetric, reflexive, and transitive.

(4.2) There exists a binary operation which can be applied to any pair a, b of elements of Ω, in this order, to produce a unique element, denoted by ab, in Ω.

(4.3) $ab=ba$; $a(bc)=(ab)c$, for all a, b, c in Ω.

(4.4) If Ω contains z such that $zx=z$ for all x in Ω, z is unique.

(4.5) If Ω contains u such that $ux=x$ for all x in Ω, u is unique, and $u\neq z$.

(4.6) If Ω contains the z in (4.4), and $ab=z$, then $a=z$ or $b=z$ (or both).

(4.7) If $ax=bx$, $x\neq z$ (if Ω contains z), then $a=b$; if Ω does not contain z, then $ax=bx$ implies $a=b$.

We need not discuss the independence of (4.1)–(4.7). The consistency is obvious from numerous instances. Note that only one binary operation is postulated.

If p, q, r are any elements $\neq z$ of Ω such that $p=qr$, we say that r *divides* p, and write $r|p$ (hence also $q|p$). In all questions of divisibility z is henceforth excluded.

If u as in (4.5) exists, and $\epsilon|u$, ϵ in Ω, ϵ is a *unit*. Hence $u=\epsilon\epsilon'$, and ϵ' is a unit; ϵ, ϵ' are *conjugate* units. Obviously u is a unit. If $\epsilon_1, \cdots, \epsilon_r$ are units, and $\epsilon_1', \cdots, \epsilon_r'$ their respective conjugates, $\epsilon_1 \cdots \epsilon_r$ and $\epsilon_1' \cdots \epsilon_r'$ are conjugate units. From (4.5), (4.7), a unit has a unique conjugate. If $x|a$ and $x|b$ imply that x is a unit, a, b are *coprime*. If $a|b$ and $b|a$, a, b are *associates*, $a\sim b$. From $a\sim b$ follows $a=\epsilon b$, ϵ unit.

An element h in Ω other than a unit such that $x|h$ only when $x\sim h$ or a unit, is *irreducible*. An irreducible element p is *prime* if $p|ab$ implies at least one of $p|a$, $p|b$. (This amounts to making all irreducibles primes—not the case in general in an algebraic integer ring.)

If $d \mid a$ and $d \mid b$ imply $d \mid g$, $g \mid a$, $g \mid b$, g is the G.C.D (by definition) of a, b.

We define Ω to be an *arithmetic* with respect to the binary operation in (4.2), and write $A\,\Omega$, if the postulates (4.8), (4.9) hold.

(4.8) If b is any element of Ω, there exist only a finite number of elements x_i of Ω different from units such that $x_i \mid b$.

(4.9) Apart from permutations of $\epsilon, p_1, \cdots, p_r$, every element b of Ω is uniquely expressible in the form $b = \epsilon p_1 \cdots p_r$, where ϵ is a unit and $p_1, \cdots, p_r$ are primes.

Rational arithmetic and the theory of algebraic numbers and ideals provide several instances of $A\,\Omega$. We have not attempted to state a minimum set of postulates sufficient for unique factorization, as we are concerned here only with the application to be made presently of the fundamental theorem to diophantine analysis. In particular, (4.9) is a consequence of the rest, which can be weakened. An exhaustive study of postulate systems leading to (4.9) has been made by Professor M. Ward in an unpublished paper.

The following consequence of the postulates will be required. If $a \mid bc$, and a, b are coprime, then $a \mid c$. For, the hypotheses are equivalent to $ad = bc$, with a, b coprime. Let $p \mid a$, where p is prime. Then $p \mid b$ or $p \mid c$. But $p \mid b$ is impossible.

5. We return to polynomial systems as defined in §3. A polynomial system will be characterized as diophantine if all integer values of the independent variables satisfying the system can be given explicitly by expressing the variables as polynomials with integer coefficients in a finite number of independent parameters ranging independently over all integers.

It will now be shown that any instance, say AVR, of VR which is an arithmetic in the sense of §4 with respect to vector multiplication as in §2 provides an infinity of polynomial diophantine systems.

The system (3.3) is purely multiplicative. Hence, since we are now operating in AVR, the method of reciprocal arrays developed in a previous paper* can be applied to obtain the complete solution of (3.3) in parametric form. The solution expresses each of the independent variables (elements of AVR) as power products of parameters in AVR. The method of arrays is applicable because it refers to any arithmetic as defined in §4. For clearness we illustrate the process by giving the first step in the proof, from which (as in the paper cited) the rest follows by mathematical induction, in the form adapted to the present discussion.

* American Journal of Mathematics, vol. 55 (1933), pp. 50–66.

For simplicity, let Greek letters denote elements of AVR for the moment. We shall find all $\alpha, \beta, \gamma, \delta$ such that

$$\alpha\beta = \gamma\delta.$$

Denote the G.C.D. of α, γ by σ. Then $\alpha = \sigma\alpha_1$, $\gamma = \sigma\gamma_1$, where α_1, γ_1 are coprime. Hence $\alpha_1\beta = \gamma_1\delta$, and therefore (by the definition of divisibility and the last of §4) $\alpha_1 | \delta$ and $\gamma_1 | \beta$. Thus $\delta = \alpha_1\delta_1$, $\beta = \gamma_1\beta_1$. With the given equation this yields $\beta_1 = \delta_1$. Denote the common value of β_1, δ_1 by τ. Then the complete solution is

$$\alpha = \alpha_1\sigma, \beta = \gamma_1\tau, \gamma = \gamma_1\sigma, \delta = \alpha_1\tau.$$

Moreover, it is sufficient to choose only such values of the parameters as are coprime.*

Consider now (3.3). The independent variables, say $\xi, \eta, \cdots, \zeta$, are in AVR. Let the $\rho, \sigma, \cdots, \tau$ be parameters ranging independently over all the elements of AVR. The method of reciprocal arrays gives the complete solution of (3.3) in the form

$$\begin{aligned} \xi &= \rho_1^{a_1} \cdots \rho_r^{a_r} \sigma_1^{b_1} \cdots \sigma_s^{b_s} \cdots \tau_1^{c_1} \cdots \tau_t^{c_t}, \\ \eta &= \rho_1^{f_1} \cdots \rho_r^{f_r} \sigma_1^{g_1} \cdots \sigma_s^{g_s} \cdots \tau_1^{h_1} \cdots \tau_t^{h_t}, \\ &\cdots \qquad \cdots \qquad \cdots \\ \zeta &= \rho_1^{k_1} \cdots \rho_r^{k_r} \sigma_1^{l_1} \cdots \sigma_s^{l_s} \cdots \tau_1^{m_1} \cdots \tau_t^{m_t}, \end{aligned}$$

where the exponents $a, b, \cdots, m$ are constant rational integers $\geqq 0$, determined by the particular forms of the power products in (3.3), and θ^0 denotes u for all θ in AVR. Since the variables $\xi, \eta, \cdots, \zeta$ are independent in AVR, and (say) $\xi = (\xi_1, \cdots, \xi_n)$, $\eta = (\eta_1, \cdots, \eta_n)$, $\cdots$, $\zeta = (\zeta_1, \cdots, \zeta_n)$, the variables $\xi_j, \eta_j, \cdots, \zeta_j (j = 1, \cdots, n)$ are independent in R. But these are precisely the independent variables in (3.4). Since each of (3.3), (3.4) implies the other, the complete solution of (3.3) yields all sets of integers (elements of R) satisfying (3.4) when $\xi_j, \eta_j, \cdots, \zeta_j$ are equated respectively to the jth coordinates in the above power products giving the general solution $\xi, \eta, \cdots, \zeta$ of (3.3). The jth coordinates in question are written down as explained in §2, and are polynomials, with rational integer coefficients, in the coordinates of the $\rho, \sigma, \cdots, \tau$. But these coordinates are parameters in R.

* In this simple example, all solutions $(\alpha, \beta, \gamma, \delta) = (\alpha_1\sigma, \gamma_1\tau, \gamma_1\sigma, \alpha_1\tau)$ are run through once only as coprime α_1, γ_1 and arbitrary σ, τ run through the elements of AVR. But in more complicated equations, the same solution may be given more than once. This however does not affect the statement that *all* solutions are given.

6. It remains to be shown that the theory is not vacuously true. For this it is sufficient to produce instances of *AVR*.

Let $R(\omega) \equiv R(\omega_1, \cdots, \omega_n)$ be an algebraic extension of R, and let $\omega_1, \cdots, \omega_n$ be a basis of $R(\omega)$ with the multiplication table

$$\omega_r \omega_s = \sum_{j=1}^{n} \phi(r, s, j)\omega_j.$$

If now $\omega_1 = u$, we may write

$$x(i) = \sum_{j=1}^{n} x(i, j)\omega_j.$$

If in particular $R(\omega)$ is the ring of all algebraic integers in an algebraic number field (relative to the rational field), $\omega_1 = 1$, and the $x(i)$ run through all integers of the field. Hence, if the field is such that unique factorization (without the introduction of ideals) holds, it is an instance of *AVR*. It can be shown conversely that any *AVR* is isomorphic with an algebraic integer ring.

If in the algebraic integer ring $R(\omega)$ there is not unique factorization, we replace the integers by the principal ideals which they generate; §4 is then applicable. But the application to (3.3) as in §5 does not then yield the solution of (3.4) *practically*, although it does *theoretically*, on account of the following elementary difficulty in the theory of ideals: Given the bases of two *general* ideals A, B to *exhibit* the basis of their product in terms of the $2n$ integers defining the bases of A, B. The use of canonical two-term bases does not remove the difficulty. If *general* in the preceding be replaced by *specific*, so that the bases of A, B are expressed in terms of *given* integers, the problem, so far as it concerns algebraic numbers, is solvable in a finite number of steps. But in that case, there is no diophantine problem (3.3) or (3.4). The general existence proof concerning a basis of $R(\omega)$ seems to lead to nothing usable for diophantine analysis.

7. The system (3.3) and its equivalent (3.4) are more restricted than is necessary. Each power product in (3.3) may be replaced by an arbitrary constant integer multiple of itself. For the discussion of (3.3) in this more general case we refer to an article in the Bulletin of the American Mathematical Society for 1933. By referring to §3 it is easily seen what the corresponding (3.4) has become: arbitrary integer coefficients are introduced.

In the papers cited, several illustrative examples of (3.3) have been given, and any desired number can be written out. In conjunction with any algebraic number field in which there is unique factorization, any such example

gives a polynomial diophantine system and its complete solution. An example is given presently.

In the second paper cited, systems not in the form (3.3) but reducible to that form by linear homogeneous substitutions on the variables, with integer coefficients, were discussed. For example, $x^2+y^2=w^2$, $x^2+y^2=w^2+t^2$, and

$$(x + y + w)^3 + t^3 + v^3 + r^3 = (t + v + r)^3 + x^3 + y^3 + z^3.$$

Such transformed (3.3) are completely solvable, and hence also the corresponding transformed (3.4).

The notation developed in §3 for the jth coordinate in any power product in VR enables us to state in concise form the system (3.4) equivalent to a given (3.3) and to write down the complete solution of a specific system (3.4) from the complete solution of the equivalent (3.3). The last is obtained directly by the algorithm of reciprocal arrays. If the explicit polynomial expressions of the coordinates are required, they are given (for a fixed base ϕ) by the first formula in §3. A simple example, where (3.3) consists of only one equation, will suffice.

The complete solution in $A\,\Omega$ of the equation

$$x^3 = ytw \tag{7.1}$$

is found by the method of arrays to be

$$\begin{aligned} x &= mabcfghpqr, \\ y &= mgh(af)^2p^3, \\ t &= mca(bg)^2q^3, \\ w &= mbf(ch)^2r^3, \end{aligned} \tag{7.2}$$

where $m, a, b, c, f, g, h, p, q, r$ are parameters ranging independently over all elements of $A\,\Omega$. We shall omit the G.C.D. conditions which may be imposed if desired, as they do not affect the generality of the solution.

By a mere change of notation (7.1) becomes (7.3) in AVR,

$$v^3(x) = v(y)v(t)v(w), \tag{7.3}$$

which is equivalent in R to the simultaneous system (corresponding to (3.4)),

$$v(x^3; j) = v(y, t, w; j) \qquad (j = 1, \cdots, n). \tag{7.4}$$

The jth coordinates written in (7.4) are homogeneous polynomials in R of degree 3, whose explicit forms can be written down by the first formula in §3. The complete solution of (7.4) is written down similarly from (7.2):

$$
\begin{aligned}
v(x;j) &= v(m, a, b, c, f, g, h, p, q, r; j),\\
v(y;j) &= v(m, g, h, a^{(2)}, f^{(2)}, p^{(3)}),\\
v(t;j) &= v(m, c, a, b^{(2)}, g^{(2)}, q^{(3)}),\\
v(w;j) &= v(m, b, f, c^{(2)}, h^{(2)}, r^{(3)}) \qquad (j = 1, \cdots, n).
\end{aligned}
$$

Thus the $4n$ independent variables in (7.4) are given parametrically in the complete solution in terms of $10n$ integer parameters.

For the complete solution of a given system (3.4) equivalent to (3.3) in an *AVR* which is algebraic of degree n it is necessary to select algebraic number fields of degree n in which there is unique factorization, and to construct the multiplication table $\omega_r\omega_s(r, s=1, \cdots, n)$ for the basis, in order to get the $\phi(r, s, j)(j=1, \cdots, n)$. For $n=2, 3, 4$ only is there sufficient knowledge extant to enable us to obtain the general ϕ. For no n is it known in all of what fields of that degree there is unique factorization; if $n=2$ any field with class number 1 may be used, but not all such fields are known; if $n=3$ there are numerous special fields known. For $n\geqq 4$, the ϕ can also be obtained for some special fields. Although there is nothing approaching generality in the available data concerning algebraic fields which is necessary for the application to diophantine analysis, nevertheless an infinity of completely solvable polynomial diophantine systems exist, and any number can be constructed from a single algebraic *AVR* alone.

As the entire subject originated in the Pythagorean equation $x^2+y^2=t^2$, we shall state the most general system equivalent to this and solvable completely in rational integers by the methods of this paper. Let d denote a non-zero rational integer, and for simplicity restrict d to have no square factor >1 (a restriction easily removed). Write $D=4d$ if $d\equiv 2$ or 3 mod 4, $D=d$ if $d\equiv 1$ mod 4; $B=-\frac{1}{4}D(D-1)$. Then B is a rational integer. Let the field generated by $d^{1/2}$ have class number 1. Then the system in question is

$$
\begin{gathered}
x_1^2 + y_1^2 - z_1^2 - w_1^2 + B(x_2^2 + y_2^2 - z_2^2 - w_2^2) = 0,\\
2(x_1x_2 + y_1y_2 - z_1z_2 - w_1w_2) + D(x_2^2 + y_2^2 - z_2^2 - w_2^2) = 0.
\end{gathered}
$$

As the complete solution of this in rational integers x_j, y_j, z_j, w_j $(j=1, 2)$ is somewhat more detailed than that of the next, which is equivalent to it, we shall conclude with the complete solution in rational integers $\xi_i, \eta_i, \lambda_i, \mu_i$ $(i=1, 2)$ of the system

$$
\begin{gathered}
\xi_1\xi_2 + B\eta_1\eta_2 = \lambda_1\lambda_2 + B\mu_1\mu_2,\\
\xi_1\eta_2 + \xi_2\eta_1 + D\eta_1\eta_2 = \lambda_1\mu_2 + \lambda_2\mu_1 + D\mu_1\mu_2.
\end{gathered}
$$

Let the $\alpha_j, \beta_j(j=1, \cdots, 4)$ be parameters ranging over all rational integers independently. Then the complete solution is

$$\xi_1 = \alpha_1\alpha_3 + B\beta_1\beta_3, \quad \eta_1 = \alpha_1\beta_3 + \alpha_3\beta_1 + D\beta_1\beta_3,$$
$$\xi_2 = \alpha_2\alpha_4 + B\beta_2\beta_4, \quad \eta_2 = \alpha_2\beta_4 + \alpha_4\beta_2 + D\beta_2\beta_4,$$
$$\lambda_1 = \alpha_1\alpha_4 + B\beta_1\beta_4, \quad \mu_1 = \alpha_1\beta_4 + \alpha_4\beta_1 + D\beta_1\beta_4,$$
$$\lambda_2 = \alpha_2\alpha_3 + B\beta_2\beta_3, \quad \mu_2 = \alpha_2\beta_3 + \alpha_3\beta_2 + D\beta_2\beta_3.$$

This follows at once, by the algorithm described, from the solution of $\alpha\beta=\gamma\delta$ in §5. Note that nothing has been proved if the class number exceeds unity.

Finally it may be stated that the number of parameters appearing in any solution obtained by the algorithm is both necessary and sufficient for the complete solution. This is a consequence of the like for any application of reciprocal arrays.

California Institute of Technology,
Pasadena, Calif.

SECTIONS OF POINT SETS*

BY

DEANE MONTGOMERY

1. Introduction

A section of a plane point set E is defined as that subset of E which contains all points of E lying on a line L. If L is a horizontal line the section is called a horizontal section and if L is a vertical line, the section is called a vertical section. It is the purpose of this paper to study the relations between E and its horizontal and vertical sections. Kuratowski and Ulam†, Sierpinski‡, and Fubini§, have considered various phases of this problem. Baire‖, Hahn¶, Kempisty** and others have considered the closely related problem of finding the relations between a function $f(x, y)$ and the functions obtained by holding x or y constant.

In order to state results in a general manner, E will be regarded as a subset of a combinatorial product space $A \times B$ where A and B are metric spaces and B is separable. Such a space is defined as the collection of all pairs of points (x, y), x being a point of A and y being a point of B. The distance between (x_1, y_1) and (x_2, y_2) is here defined to be $[(x_1x_2)^2+(y_1y_2)^2]^{1/2}$.†† The plane is a special case of such a space in which A and B are straight lines, and all the results of this paper apply to the plane and also to an $(m+n)$-dimensional euclidean space considered as the product of an m-dimensional and an n-dimensional euclidean space.

Because $A \times B$ is analogous to the plane, the subset of points (x, y) such that $x=a$ is called a vertical section of $A \times B$ and is denoted by $a \times B$ or $(x=a)$; similarly the subset of points (x, y) such that $y=b$ is called a horizontal section of $A \times B$ and is denoted by $A \times b$ or $(y=b)$. If E is any subset of $A \times B$ the set $E\cdot(x=a)$ is called a vertical section of E and the set $E\cdot(y=b)$ is called a horizontal section of E.

---

* Presented to the Society, November 25, 1932; received by the editors February 1, 1933.

† Fundamenta Mathematicae, vol. 19, p. 247; see also an article by Kuratowski in vol. 17, p. 275.

‡ Fundamenta Mathematicae, vol. 1, p. 112.

§ Rendiconti della Reale Accademia dei Lincei, (5), vol. 16, I. For a statement of Fubini's theorem see also Carathéodory, *Vorlesungen über Reelle Funktionen*, 1927, p. 621.

‖ Annali di Matematica, 1899, p. 1.

¶ Mathematische Zeitschrift, 1919, p. 306.

** Fundamenta Mathematicae, vol. 14, p. 237, and vol. 19, p. 184.

†† If p and q are any two points of a metric space, (pq) denotes the distance between p and q.

If E is closed, all horizontal and vertical sections of E are closed, and if E is open, its horizontal and vertical sections are open (relative to the sections of $A \times B$ which contain them). A similar proposition is true for sets F or O of type α. Converse propositions are not true. The plane set of points $(1/n, 1/n)$, where n takes all integral values, is such that each of its horizontal and vertical sections contains at most one point and is therefore closed. The point $(0, 0)$ is a limit point of the set which is not in the set. Sierpinski* has constructed a plane set every section of which (not merely horizontal and vertical sections) contains at most two points and which is non-measurable in the Lebesgue sense. This example shows that the fact that every horizontal and vertical section of E is of type α is not a sufficient condition that E be of type α, and that in order to obtain such a sufficient condition, further restrictions on the sections or on the relations between them must be imposed. By restricting the vertical sections to a type of set called I-set (or the complement of such a set) sufficient conditions may be obtained that a set be of various types. This is done in §3. Necessary and sufficient conditions that sets with restricted vertical sections be of class α are given in §6. Uses of sets called gratings are considered in §7. Theorems are given which show that boundaries of sets with certain kinds of sections lie on sets of lines of the first category. The results are applied in §8 to prove a theorem of Baire concerning functions of two variables continuous in each of them and to obtain a result regarding Kempisty's generalization of this theorem.

2. Horizontal sections of class M

The following definitions will be useful.

Definition 1. *If the inner points of a set are dense on the set, the set is called an I-set.*

A set may have this property with respect to $A \times B$ or it may be a subset of a section of $A \times B$ and have this property with respect to the section, this latter being the case which will most often arise.

Definition 2. *Given a point* (a, b), *the set of points* (a, y) *where* $(by) < r$, r *a positive number, is called an open vertical interval of center* (a, b) *and radius* r.

A closed vertical interval is defined in the same way except that $(by) \leqq r$. Closed and open horizontal intervals may also be defined. A vertical interval might also be defined as $a \times S$ where S is a sphere in B of center b and radius r.

* See the previous reference.

DEFINITION 3. *If a set G lies on a horizontal section, $G(r)$ is the set of points of open vertical intervals of radii r and centers at the points of G.*

DEFINITION 4. *If $\mathfrak{M}$ is a family* of point sets M lying on horizontal sections of $A\times B$, $\mathfrak{M}(r)$ is the family of all point sets $M(r)$ for r ranging over all positive numbers.*

If $\mathfrak{M}$ is a family of point sets, $\mathfrak{M}_\sigma(\mathfrak{M}_\delta)$ denote, as is conventional, the families composed of all possible sums (products) of an enumerable number of sets of $\mathfrak{M}$.

A set of vertical sections K is said to be everywhere dense if the set of points $[a]$, such that $(x=a)$ is in K, is everywhere dense in A. In a similar manner other point-set properties, for example the property of being in the first or the second category, are ascribed to sets of vertical sections and to sets of horizontal sections as well.

By the projection of a point (x, y) on $(y=b)$ is meant the point (x, b). The projection of a set of points E on $(y=b)$ is the set of points formed by projecting all the points of E on $(y=b)$.

THEOREM 1. *If each vertical section of E is an open set whose complement is an I-set† and horizontal sections of E belong to $\mathfrak{M}$, then E belongs to the family $[\mathfrak{M}_\delta(r)]_\sigma$.*

Since B is separable there exists an enumerable everywhere dense set $(y=b_i)$ of horizontal sections of $A\times B$. Let r_i be a sequence of positive numbers approaching 0 as a limit. Let K_j be the projection of $E\cdot(y=b_j)$ on $(y=b_i)$. Let A_{im} be the product of $E\cdot(y=b_i)$ and all sets K_j such that $(b_jb_i)<r_m$. The set $A_{im}(r_m)$ is a subset of E. To prove this, suppose that $A_{im}(r_m)$ contains a point (a, b) of CE. The point (a, b_i) is in $A_{im}(r_m)$ and also all points of the open vertical interval of radius r_m and center (a, b_i) are in $A_{im}(r_m)$. Since (a, b) lies in this open vertical interval, there is by hypothesis some inner point‡ (a, c) of CE in this interval. There is then an ϵ such that if $(cy)<\epsilon$, (a, y) is in CE. Since the b_i's are everywhere dense in B there is some b_i, say b_n, such that $(cb_n)<\epsilon$. The point (a, b_n) is then in CE. This b_n may be so chosen that $(b_ib_n)<r_m$. The set $E\cdot(y=b_n)$ does not con-

* It is here supposed that if E consisting of points (x, b) is in $\mathfrak{M}$, the set of points (x, c), where x has the same range as in E and c is any point of B, is also in $\mathfrak{M}$. This restriction is made for convenience. It is necessary for the proofs of some of the following theorems in which use is made of the projections of sets from one horizontal section to another.

† A more explicit statement is as follows: If V is any vertical section of $A\times B$, $V\cdot E$ is open in V and $V-E$ is an I-set in V, etc. Language similar to that in the hypothesis of Theorem 1 will be used throughout, with the meaning given in this note.

‡ With respect to the section $(x=a)$.

tain (a, b_n) and K_n will not contain (a, b_i). Therefore (a, b_i) will not be in $A_{im}(r_m)$. From this contradiction, it follows that $A_{im}(r_m)$ is in E.

To prove that every point of E is in some $A_{im}(r_m)$, let (c, d) be any point of E. There is an ϵ such that if $(dy) < \epsilon$, (c, y) is in E. Choose r_m and b_i such that $b_i d < r_m < \epsilon/2$. Every point (c, y) of the open vertical interval of center (c, b_i) and radius r_m is in E. For since $(b_i y) < \epsilon/2$ and $(b_i d) < \epsilon/2$, it follows from the triangle axiom that $(dy) < \epsilon$. Therefore the set K_j contains (a, b_i) if $(b_i b_j) < r_m$, and $A_{im}(r_m)$ contains all points (c, y) such that $(b_i y) < r_m$. It must then contain (c, d) since $(b_i d) < r_m$.

Thus $E = \sum_{im} A_{im}(r_m)$, which proves the theorem.

THEOREM 2. *If vertical sections of E are closed I-sets and horizontal sections of E belong to the family* $\mathfrak{M}$, *then E belongs to the family* $[\mathfrak{M}_\sigma(r)]_{\sigma\delta}$.

Let $(y = b_i)$ be an enumerable everywhere dense set of horizontal sections of $A \times B$ and let (r_i) be a sequence of positive numbers approaching 0. Let K_j be the projection of $E \cdot (y = b_j)$ on $(y = b_i)$. Let A_{im} be the sum of $E \cdot (y = b_i)$ and all sets K_j such that $(b_i b_j) < r_m$. The set $E_m = \sum_i A_{im}(r_m)$ is a member of the family $[\mathfrak{M}_\sigma(r)]_\sigma$. The set E_m contains E, for all m. To prove this let (c, d) be any point of E. By hypothesis there is for each r_m a vertical interval V_e, of radius e, which contains only points of E and for every point (c, y) of which $(dy) < r_m$. There is some $(y = b_i)$ which cuts V_e in a point (c, b_i). The point (c, b_i) is then in E. The set $A_{im}(r_m)$ contains (c, d) because $(db_i) < r_m$.

The set E is therefore included in $\prod_m E_m$. To prove that $E = \prod_m E_m$, let (a, b) be any point of CE. There is an ϵ such that if $(by) < \epsilon$, (a, y) is in CE. In order that (a, b) be in $A_{im}(r_m)$ it is necessary that $(bb_i) < r_m$. Choose $r_m < \epsilon/2$. If $(bb_i) < r_m$, and $(b_i y) < r_m$, it follows that $(by) < \epsilon$, and consequently (a, y) is in CE. Therefore if $(bb_i) < r_m$, $A_{im}(r_m)$ cannot contain (a, b) because all points on the open vertical interval of center (a, b_i) and radius r_m are in CE. As has been mentioned, $A_{im}(r_m)$ cannot contain (a, b) if $(bb_i) \geqq r_m$. Therefore for r_m chosen as it has been, (a, b) is not in E_m and consequently not in $\prod_m E_m$. It follows that $E = \prod_m E_m$, which completes the proof of the theorem.

3. HORIZONTAL SECTIONS OF THE TYPE α

LEMMA 1. *Let G be a set lying on a horizontal section of* $A \times B$. *If G is an* $O_\alpha (\alpha \geqq 0)$ *or an* $F_\alpha (\alpha > 0)$, $G(r)$ *is an* O_α *or an* F_α.* *If G is analytic,* $G(r)$ *is analytic.*

The first part of the lemma is true for sets O_0 and sets O_1, and can be shown to be true for sets O_α by transfinite induction. The latter part of the lemma may be readily proved from the definition of an analytic set.

* For a discussion of sets F_α and O_α, see de la Vallée Poussin, *Intégrales de Lebesgue*, p. 132.

THEOREM 3. *If the horizontal sections of E are O_α's and the vertical sections are closed I-sets, then E is an $F_{\alpha+1}$.*

This follows from Theorem 2. $\mathfrak{M}$ is here the family of O_α's in horizontal sections of $A\times B$. $\mathfrak{M}_\sigma$ is the same family and by the preceding lemma $\mathfrak{M}_\sigma(r)$ is a family of O_α's in $A\times B$, as is $[\mathfrak{M}_\sigma(r)]_\sigma$. Therefore $[\mathfrak{M}_\sigma(r)]_{\sigma\delta}$ is a family of $F_{\alpha+1}$'s in $A\times B$ and E must be an $F_{\alpha+1}$.

By taking complements the following is proved:

THEOREM 4. *If the horizontal sections of E are F_α's and the vertical sections are open sets whose complements are I-sets, then E is an $O_{\alpha+1}$.*

That the change in classification mentioned in Theorems 3 and 4 actually may occur, is shown by the following plane set. On the line $y=x$, take a set E which is an $O_{\alpha+1}(\alpha\geqq 1)$* and an O of no lower class.† Then $E=\sum_i A_i$ where A_i is an F_α at most. At each point of A_i erect a vertical interval of length $1/i$, closed at the end touching $y=x$ and open at the other end, and denote the set thus obtained by H. Horizontal sections of H are F_α's. This can be seen as follows. If L is any horizontal line and p any point on this line above the line $y=x$, vertical intervals from only a finite number of the sets A_i can cut L to the left of, or at, p. Therefore the points of H on L to the left of or at p must form an F_α. This is true however close p may be to $y=x$. Let p_n be a sequence of points on L above $y=x$, approaching $y=x$. Let E_n be the points of $(CH)\cdot L$ to the left of or at p_n. Then E_n is an O_α, and $\sum E_n$ is an O_α. Therefore the points of CH on L to the left of $y=x$ form an O_α and consequently the points of H on L form an F_α. Since $\alpha\geqq 1$, it does not matter whether or not the intersection of L and $y=x$ is in H. Although horizontal sections of H are F_α's, the set H itself must be at least an $O_{\alpha+1}$ since $y=x$ cuts it in an $O_{\alpha+1}$.

Denote by R that part of the complement of H which lies on or above $y=x$. Horizontal sections of R are O_α's but R itself is an $F_{\alpha+1}$ at least, since $y=x$ cuts it in an $F_{\alpha+1}$. By Theorem 3, R is an $F_{\alpha+1}$ at most. This example shows that under the hypothesis of Theorem 3, it is impossible to draw a stronger conclusion on the F classification of E than the one there given.

THEOREM 5. *If the horizontal sections of E are O_α's and the vertical sections are open sets whose complements are I-sets, then E is an $O_{\alpha+2}$.*

This follows from Theorem 1. The family $\mathfrak{M}$ is here the family of sets O_α on horizontal sections of $A\times B$. $\mathfrak{M}_\delta$ is then the family of sets $F_{\alpha+1}$ on hori-

* A single open vertical interval furnishes an example in case $\alpha=0$.

† For a proof of the existence of functions of all classes (which proves the existence of sets of all classes) see de la Vallée Poussin, *Intégrales de Lebesgue*, p. 145 ff.

zontal sections. $\mathfrak{M}_\delta(r)$ contains only sets $F_{\alpha+1}$ in $A\times B$ by Lemma 1. Therefore $[\mathfrak{M}_\delta(r)]_\sigma$ contains only sets $O_{\alpha+2}$.

By taking complements there is proved

THEOREM 6. *If the horizontal sections of E are F_α's and the vertical sections are closed I-sets, then E is an $F_{\alpha+2}$.*

Whether or not the classification may actually be increased by two under the hypothesis of Theorems 5 and 6 is an open question. That an advance of one may occur is shown by the following plane set. Construct the set H as in the preceding example, except that the vertical intervals are now to be closed instead of half closed. Horizontal sections of this set are F_α's as before and the set itself must be an $O_{\alpha+1}$ exactly as H was.

The two following theorems may be proved directly or as a result of the preceding theorems on sets O_α and F_α.

THEOREM 7. *If horizontal sections of E are A_α's *and vertical sections of E are all closed I-sets or all open sets whose complements are I-sets, then E is an $A_{\alpha+2}$.*

4. ANALYTIC OR MEASURABLE HORIZONTAL SECTIONS

If $\mathfrak{M}$ is the family of analytic sets, $\mathfrak{M}_\sigma$ and $\mathfrak{M}_\delta$ are the same family, from which we have the following theorem.

THEOREM 8. *If horizontal sections of E are analytic and vertical sections of E are all closed I-sets or all open sets whose complements are I-sets, E is analytic.*

If there is a theory of measure in the space under consideration, as for example in the plane, a theorem similar to Theorem 8 is true for measurable sets.

5. THE SET E_e

Let E_e denote the subset of E each point of which lies on a closed vertical interval of radius exactly e, which contains only points of E.

For the two theorems of this section the conditions on A and B are that they are metric, that every closed sphere in B is compact and that inner points of a closed sphere in B are dense on the sphere.

THEOREM 9. *If horizontal and vertical sections of E are closed, E_e is closed.*

Let (a_n, b_n) be a sequence of points of E_e converging to a limit point (a, b). Each (a_n, b_n) lies on a closed vertical interval, V_e^n, containing only points of E, of radius e, and of center (a_n, c_n). An infinite number of points b_n are such that $(bb_n)<\epsilon$, where ϵ is any positive number. The fact that $(b_nc_n)\leqq e$ for all n

* For a discussion of sets A_α, see de la Vallée Poussin, *Intégrales de Lebesgue*, p. 135.

implies that for an infinite number of points c_n, $(bc_n) < e + \epsilon$, and by hypothesis, these points have some limit point c such that $(bc) \leqq e$. Let V_e be the closed vertical interval of center (a, c) and radius e. Let (a, y) be any point such that $(cy) < e$, that is, any point on the interior of the vertical interval V_e. It will now be shown that there is an n for which (a_n, y) is in E and $(a_n a) < \eta$, where η is any positive number. Consider the n's for which $(a_n a) < \eta$ and select from this group an n such that $(c_n c) < e - (cy)$, that is, such that $(c_n c) + (cy) < e$. It follows that $(c_n y) < e$ and therefore the point (a_n, y) is in E. On the horizontal section $(y = y)$ there is thus a sequence of points of E approaching (a, y), and since horizontal sections of E are closed, (a, y) must be in E. Therefore, all inner points of the vertical interval V_e are in E. Because of the hypothesis on the space B and the fact that vertical sections of E are closed, it follows that every point of V_e is in E. Since $(bc) \leqq e$, the point (a, b) is in V_e, which proves that every limit point of E_e is in E_e.

For Theorem 10, B is required to have the further property that any point p of B on a closed sphere of radius $> r$ is on a sphere of radius exactly r, which is contained in the first sphere.

THEOREM 10. *If horizontal and vertical sections of E are closed and each point of E lies on a closed vertical interval containing only points of e, then E is an F_σ.*

Let (r_i) be a sequence of positive numbers approaching 0. Any point in E_r is in E_{r_i} if $r_i \leqq r$. Hence $E = \sum E_{r_i}$ and since E_{r_i} is closed, E is an F_σ.

6. UNIFORMITY PROPERTIES

DEFINITION 5. *A point p is said to be a point of uniformity of E if, for some open sphere S in $A \times B$ of center p and for some e, $E \cdot S = E_e \cdot S$.*

DEFINITION 6. *A point p is said to be a point of uniform separation of E if it is a point of uniformity of CE.*

The points of uniformity of E form an open set.

For the theorems of this section, A and B are metric separable spaces. In addition it is required that every sphere in B be totally limited* and that B have the property stated just before Theorem 10.

THEOREM 11. *If the vertical sections of E are open, a necessary and sufficient condition that E be an O_α is that the set of points of non-uniform separation of E in E be an O_α and that horizontal sections of E be O_α's.*

The necessity will first be demonstrated. If E is an O_α its horizontal sections are O_α's. Let N denote the set of points of non-uniform separation of E.

* See Hausdorff, *Mengenlehre*, 1927, p. 108.

Since N is closed, it follows that $E\cdot N$ is an O_α if α is greater than 0. For $\alpha=0$, the necessity is obvious.

The sufficiency will now be demonstrated. Let p be any point of E which is a point of uniform separation of E. For some open sphere S of center p and some e, $S\cdot(CE)_e=S\cdot(CE)$. Let r_i be a sequence of positive numbers approaching 0 and let $(y=b_i)$ be an enumerable everywhere dense set of horizontal sections of $A\times B$. Let r_m be a fixed element of the sequence r_i and let ϵ be the smaller of the two numbers $e/4$ and $r_m/4$. Let S_1 be a sphere with the same center as that of S and with radius $2e$ larger than that of S. Because the projection of S_1 in B is a sphere in B it is totally limited in B by hypothesis. There exists then a finite set of horizontal sections $(y=h_i)$ such that every point of S_1 is a distance less than ϵ from some one of them. Let K_j be the projection of $S\cdot E\cdot(y=h_j)$ on $(y=b_i)$. Denote by A_{im} the product of $S\cdot E\cdot(y=b_i)$ and all K_j's such that $(h_jb_i)<r_m$.

In order to show that $A_{im}(r_m/2)$ is in $S\cdot E$, let (a, y) be any point of $A_{im}(r_m/2)$ and suppose that (a, y) is in $C(S\cdot E)$. It is then on a closed vertical interval V_e, of radius e, containing only points of $C(S\cdot E)$. By the hypothesis on the space B, the point (a, y) is also on a vertical interval V_ϵ, of radius exactly ϵ, V_ϵ being contained in V_e. The interval V_ϵ contains only points of $C(S\cdot E)$. Let (a, b) be the center of V_ϵ. Since (a, b) must be in S_1 there is some h_n such that $(bh_n)<\epsilon$. The point (a, h_n) is therefore in $C(S\cdot E)$. From the relations $(b_iy)<r_m/2$ and $(by)<\epsilon$, it follows that $(bb_i)<\frac{3}{4}r_m$. Since $(bh_n)<\epsilon$, it follows that $(h_nb_i)<\frac{3}{4}r_m+\epsilon<r_m$. The point (a, h_n), being in $C(S\cdot E)$, cannot be in $S\cdot E\cdot(y=h_n)$. Therefore (a, b_i) is not in K_n nor in A_{im}. Neither (a, b_i) nor (a, y) can then be in $A_{im}(r_m/2)$. From this contradiction it follows that $A_{im}(r_m/2)$ is in $S\cdot E$.

The proof showing that each point of $S\cdot E$ is in some $A_{im}(r_m/2)$ is analogous to the proof of a similar proposition given in the demonstration of Theorem 1, and will not be repeated here. Assuming this to be proved, $S\cdot E=\sum_{im}A_{im}(r_m/2)$. The set A_{im} is a finite product of O_α's and must be an O_α. Each $A_{im}(r_m/2)$ must then be an O_α, and therefore $S\cdot E$ is an O_α.

Every point of uniform separation of E is therefore the center of an open sphere S such that $S\cdot E$ is an O_α. By Lindelöf's* theorem an enumerable set (S_i) of such spheres cover the points of uniform separation of E in E. The set N of points of non-uniform separation of E in E is an O_α by hypothesis. Therefore $E=\sum_i S_i\cdot E+N$ is an O_α.

---

* This holds since $A\times B$ is separable when A and B are. See the previously cited paper by Kuratowski and Ulam.

THEOREM 12. *If the vertical sections of E are closed, a necessary and sufficient condition that E be an* $F_\alpha(\alpha>0)$ *is that the set of points of non-uniformity of E in E be an* F_α *and that horizontal sections of E be* F_α*'s.*

The necessity will first be demonstrated. Horizontal sections of E are F_α's since they are the products of E and horizontal sections of the space $A\times B$. The set N of points of non-uniformity of E is closed and the product of N and E must be an F_α.

The sufficiency will now be shown. Let p be any point of uniformity of E. It is a point of uniform separation of CE and for some sphere S of center p, $(CE)\cdot S$ is an O_α by Theorem 11. As before, an enumerable number of such spheres, S_i, cover the points of uniformity of E. Since the points of CE in $\sum_i S_i$ are an O_α, the points of E in $\sum_i S_i$ form an F_α. The set N of points of non-uniformity of E in E form an F_α by hypothesis. Since $E=\sum_i E\cdot S_i+N$, E is an F_α.

The theorem is not true for $\alpha=0$.

7. GRATINGS AND CATEGORICITY

In this section A and B are to be metric, separable, locally compact* spaces. In such spaces the complement of a set of the first category is of the second category, and open sets are of the second category. These propositions may be proved by a method similar to the method used for proving them in euclidean space. It is necessary to make use of the fact that every monotonic decreasing sequence of non-null compact spheres has a non-null product.†

DEFINITION 7. *If a horizontal section* $(y=b)$ *contains a set H of the second category* [*in* $(y=b)$], *such that* $H(r)$ *is in E for some r,* $(y=b)$ *is said to have property C with respect to E.*

DEFINITION 8. *If a horizontal section* $(y=b)$ *contains a set H, in and everywhere dense in a set O* [*in and open in* $(y=b)$] *such that* $H(r)$ *is in E for some r,* $(y=b)$ *is said to have property D with respect to E.*

The set $H(r)$ of Definition 8 is called a grating. A point p is said to be *within* the grating if p is in $O(r)$. A point p is *on* the grating if it is in $H(r)$.‡ Property C implies property D since a set of the second category must contain a subset everywhere dense in some open set.

LEMMA 2. *If* $\mathfrak{K}$, *a set of vertical sections, is of the second category and if each* $K\epsilon\mathfrak{K}$ *contains a vertical interval including only points of E, then some horizontal section has properties C and D with respect to E; furthermore the center of one of the vertical intervals is on the grating* (*of property D*).

* For a definition of this term see Fréchet, *Les Espaces Abstraits*, p. 223.

† See Banach, *Théorie des Opérations Linéaires*, pp. 13 and 14.

‡ If p is *on* a grating it is *within* the grating.

Let K_{3e} be the set of vertical sections containing vertical intervals of E of radii $>3e$, where e has been so chosen that K_{3e} is of the second category. Let V_{3e} denote an individual one of these vertical intervals of radius $>3e$ and let $\sum V_{3e}$ denote the points in all such intervals. From each V_{3e} form a vertical interval V_e of radius exactly e with the same center as V_{3e}. Each interval V_e consists entirely of points of E. Let (b_i) be an enumerable set of points everywhere dense in B, and let B_i be a sphere in B of center b_i and radius $2e$. Let A_i be the subset of A such that for each $a\epsilon A_i$, there is some V_e and a corresponding V_{3e} for which $V_e < a\times B_i < V_{3e}$.

It will now be shown that for each V_e and corresponding V_{3e}, of center (a, b), there is an i such that $V_e < a\times B_i < V_{3e}$, and consequently that $\sum A_i$ is the set in which the sections of K_{3e} cut A. In order to do this, choose i such that $(bb_i) < e$. For each point (a, y) of V_e, $(by) < e$. By the triangle axiom, $(b_iy) < 2e$. This shows that the vertical interval $a\times B_i$ of center (a, b_i) and radius $2e$ includes V_e. It will now be shown that $a\times B_i < V_{3e}$. If (a, y) is any point of $a\times B_i$, $(b_iy) < 2e$. Since $(bb_i) < e$, it follows from the triangle axiom that $(by) < 3e$ which is the condition that (a, y) be in V_{3e}.

Since $\sum A_i$ is the set in which the sections of K_{3e} cut A, it follows that $\sum A_i$ is of the second category and, consequently, that some particular A_i, say A_n, is of the second category. The section $(y=b_n)$ has property C with respect to E, for the set $A_n\times b_n$ is of the second category in $(y=b_n)$, and each point of $A_n\times b_n$ is the center of a vertical interval of radius $2e$ which includes only points of E. The section $(y=b_n)$ must then also have property D with respect to E. Since each of these vertical intervals contains an interval V_e it must contain the center of this interval V_e which is the center of the corresponding original interval V_{3e}. Therefore the center of one of the original intervals is on each vertical interval of the grating.

It is evident that when A and B have similar properties, the parts played by horizontal and vertical sections in any theorem may be interchanged.

Definition 9. *A point is said to be of the second category with respect to E if every neighborhood of the point contains a subset of E of the second category.*

The set of all points of the second category with respect to E is denoted by E_{sc}. The set E_{sc} is closed.

A necessary and sufficient condition that E be of the second category is that E_{sc} be of the second category. This implies that if E is of the second category, it must be of the second category at a set everywhere dense in an open set, and since E_{sc} is closed it must be of the second category at each point of an open set.†

† See Banach, *Théorie des Opérations Linéaires*, p. 13 and the reference there given.

THEOREM 13. *If vertical sections of R are I-sets and horizontal sections of T are I-sets, and* $R\cdot T=0$, *then* $R'\cdot T+R\cdot T'$ *is of the first category in* $A\times B$.

It is sufficient to show that $R\cdot T'$ is of the first category in $A\times B$. Assume that $R\cdot T'$ is of the second category in $A\times B$. It must then be of the second category at every point of a set O, open in $A\times B$, which implies that R and T are both dense in O. It follows from the hypothesis, that on each vertical section containing a point of $R\cdot T'\cdot O$, there is a vertical interval V containing only points of R and such that V is in O. Since the vertical sections containing points of $R\cdot T'\cdot O$ form a set of the second category†, Lemma 2 may be applied. By this lemma, there is a horizontal section L containing a set H everywhere dense in O^* (O^* a set open in L) such that $H(r)$ is in R. The set $H(r)$ is also in O since the intervals V from which it is constructed are in O. Suppose there is a point p of T within $H(r)$ and let the horizontal section containing p be L^*. The section L^* must contain an inner point (with respect to L^*) of T which lies in $O^*(r)$. But this is impossible because $L^*\cdot H(r)$ is dense in $L^*\cdot O^*(r)$ and $H(r)$ contains only points of R. There can be, then, no points of T within $H(r)$, but this is a contradiction since T must be dense in O. Therefore $R\cdot T'$ is of the first category in $A\times B$.

COROLLARY 1. *If vertical sections of E are I-sets and horizontal sections of CE are I-sets, then* $E'\cdot(CE)+E\cdot(CE)'$ *is of the first category in* $A\times B$.

This follows immediately from the theorem and the fact that $E\cdot(CE)=0$.

COROLLARY 2. *If vertical sections of E are I-sets and horizontal sections of CE are I-sets, there is an inner point of either E or CE in every set O, open in* $A\times B$.

This follows from Corollary 1. Points not belonging to $E'\cdot(CE)+E\cdot(CE)'$ are inner points either of E or of CE and this set is everywhere dense in $A\times B$.

THEOREM 14. *If horizontal and vertical sections of E are I-sets and horizontal sections of CE are I-sets, then E is an I-set in* $A\times B$.

Let p be any point of E lying on a horizontal section L and let O be any open set in $A\times B$ containing p. It is necessary to show that O contains an inner point of E. By hypothesis $L\cdot E$ contains a set O^* open in L. The set O^* is of the second category in L, and each vertical section K cutting O^* must contain a vertical interval V including only points of E and lying in O. From these V's, there may be formed a grating $H(r)$ containing only points of E and contained in O. No point of CE can be within this grating because hori-

† If this were not true, $R\cdot T'\cdot O$ would be of the first category in $A\times B$.

zontal sections of CE are I-sets. The argument is similar to the one in Theorem 13 and will not be repeated.

Kuratowski and Ulam have a theorem similar to the following:

THEOREM 15. *If E is a set whose horizontal sections are I-sets, and O is an open set in $A\times B$, a necessary and sufficient condition that $E\cdot O$ be dense in O, is that the vertical sections L for which $L\cdot E\cdot O$ is not dense in $L\cdot O$ form a set $\mathcal{L}$ of the first category.*

The sufficiency of the condition follows from the fact that if vertical sections K, such that $K\cdot E\cdot O$ is dense in $K\cdot O$, form a set complementary to a set of the first category, they are everywhere dense and therefore the points in $K\cdot E\cdot O$, considering all K, must be dense in O.

In order to prove the necessity, let $E\cdot O$ be dense in O and suppose the set $\mathcal{L}$ to be of the second category. On each L there is a vertical interval which is in O and which contains no points of E, that is, it contains only points of CE. By Lemma 2, CE must contain a grating which is in O. But this grating can have within it no point of E since horizontal sections of E are I-sets. This contradicts the hypothesis that E is dense in O, and the theorem is proved.

THEOREM 16. *If vertical sections of R are open, horizontal sections of T are I-sets and $R\cdot T=0$, then $R\cdot T'$ lies on a set $\mathcal{K}$ of vertical sections, which is of the first category.*

Suppose that $\mathcal{K}$ is of the second category. Each $K\epsilon\mathcal{K}$ contains a vertical interval with a point of $R\cdot T'$ as center and containing only points of R. By Lemma 2, there is a grating, composed of points of R, containing a point of $R\cdot T'$ on its interior. This is impossible because horizontal sections of T are I-sets.

COROLLARY 3. *If vertical sections of E are closed and horizontal sections of E are I-sets, then $(CE)\cdot E'$ lies on a set $\mathcal{K}$, of vertical sections, which is of the first category.*

This corollary may be proved by replacing R and T by CE and E in Theorem 16.

If there exists a set $\mathcal{K}$ of vertical sections, and a set $\mathcal{L}$ of horizontal sections, so that every point of E lies either on a member of $\mathcal{K}$ or a member of $\mathcal{L}$, the set E is said to lie on the set $\mathcal{K}$ *plus* the set $\mathcal{L}$. This language is used to distinguish this case from the case in which every point of E lies both on a member of $\mathcal{K}$ and on a member of $\mathcal{L}$; in this latter case E is said to lie on the set $\mathcal{K}$ *and* on the set $\mathcal{L}$.

Corollary 4. *If vertical sections of R are open, horizontal sections of T are open and* $R\cdot T=0$, *then* $R'\cdot T+R\cdot T'$ *lies on a set* $\mathcal{K}$ *of vertical sections plus a set* $\mathcal{L}$ *of horizontal sections, both* $\mathcal{K}$ *and* $\mathcal{L}$ *being of the first category.*

This is a slightly stronger conclusion than that of Theorem 13, made possible by the stronger hypothesis given here.

Corollary 5. *If vertical and horizontal sections of both R and T are open, and* $R\cdot T=0$, *then* $R'\cdot T+R\cdot T'$ *lies on a set* $\mathcal{K}$ *of vertical sections and a set* $\mathcal{L}$ *of horizontal sections, both* $\mathcal{K}$ *and* $\mathcal{L}$ *being of the first category.*

In Corollary 5, the projection of $R'\cdot T+R\cdot T'$ on any horizontal or vertical section must be of the first category in the section. This is not necessarily true in Corollary 4.

It will be assumed in the following theorem that the space A is dense in itself in order that the set $\overline{E}$ there considered may be perfect. It will also be assumed that A and B have the properties necessary to apply Theorem 10.

A point of closure of a set E is a point in some neighborhood of which E is closed.

Theorem 17. *If horizontal and vertical sections of E are closed and each point of E lies on a closed horizontal interval containing only points of E, points of closure of E in E are dense on E.*

By Theorem 10, E is an F_σ. It must then be an F_σ in $\overline{E}=E+E'$.† It is necessary to prove that $\overline{E}-E$ is nowhere dense in $\overline{E}$ or, in other words, that limit points of E not in E are not dense on E. It will be shown that $\overline{E}-E$ is of the first category. By Corollary 3, $\overline{E}-E$ (which is the same as $(CE)\cdot E'$) lies on a set $\mathcal{K}$, of vertical sections, of the first category. Let A_0 be the set of points in which the sections of $\mathcal{K}$ cut A. The set $A_0=\sum A_i$ where each A_i is nowhere dense in A. Let R_i be the points of $\overline{E}-E$ lying on those sections, of the set $\mathcal{K}$, which cut A in A_i. If R_i were dense on some portion of $\overline{E}$, it would have to have as a limit point every point of some horizontal interval. This follows from the hypothesis that every point of E lies on a horizontal interval. This is impossible since the projection, A_i, of R_i on A, would then be everywhere dense in some open set in A. Therefore R_i is nowhere dense in $\overline{E}$, and $\overline{E}-E=\sum R_i$ is of the first category in $\overline{E}$. It follows that $\overline{E}-E$ is nowhere dense in $\overline{E}$, for $\overline{E}-E$ is a G_δ and if a G_δ is of the first category, it is nowhere dense.‡

---

† In this particular case $\overline{E}=E'$.

‡ For example see Blue, Mathematische Annalen, vol. 102, p. 627, in the proof of Theorem 1.

8. Applications

The spaces A and B are restricted here in the same manner as in §7. An interesting application of Corollary 5 is in the proof of the following result of Baire†:

If $f(x, y)$ is a real-valued function defined on the space $A\times B$ and is continuous in each of the variables separately, points of discontinuity of $f(x, y)$ lie on a set of vertical sections *and* a set of horizontal sections, both sets of sections being of the first category.

Let (r_i) be the set of rational numbers. Let R_i be the points of $A\times B$ at which $f(x, y)>r_i$ and let T_j be the set at which $f(x, y)<r_j$. From the properties of continuous functions, R_i and T_j have open horizontal and vertical sections and are disjoined. It follows that $R_i'\cdot T_j+R_i\cdot T_j'$ lies on a set of horizontal *and* a set of vertical sections of the first category, and the same is true of $\sum_{ij}(R_i'\cdot T_j+R_i\cdot T_j')$, the sum being taken only over pairs of i and j for which $r_j<r_i$. The points of this sum are the points of discontinuity of $f(x, y)$.

Applying Corollary 4 in the same way to a function upper semi-continuous in one variable and lower semi-continuous in the other, it may be shown that the set of discontinuities, E, of such a function lies on a set of horizontal sections *plus* a set of vertical sections of the first category. The set E is of the first category, but not all sets of the first category lie on a set of horizontal *plus* a set of vertical sections of the first category. This conclusion therefore contains a result not given by Kempisty.‡

---

† Acta Mathematica, 1899, p. 94.

‡ Fundamenta Mathematicae, vol. 14, p. 237.

University of Iowa,
Iowa City, Iowa

INVARIANTS OF PFAFFIAN SYSTEMS*

BY

MABEL GRIFFIN

1. **Introduction.** The developments in this paper are based on a series of invariant sets of forms associated with a given pfaffian system. These forms are obtained by exterior multiplication of the given pfaffians and their derived forms. The first set Ω^i of r forms, obtained in turn by multiplying the product of all the given pfaffians by each of the derived forms, has been employed by Cartan. The vanishing of this set is a necessary and sufficient condition that the system be passive. The present paper interprets the vanishing of the second set in the light of the notion of a *primitive system*, i.e., one whose derived system is the given set of equations.

Associated with any pfaffian system are invariant pfaffian systems, formed by equating to zero the linear factors common to any set of forms, $\Omega^{ij\cdots k}$. Any arithmetical invariant of one of the latter systems is invariant for the former; such invariants are the number of equations in the system, the class, the species, etc.

Another arithmetical invariant arises from the fact that every system possesses a primitive system.

The importance of arithmetical invariants for pfaffian systems lies in their usefulness in determining the non-equivalence of pfaffian systems. Riquier† has given methods which may be applied to determine whether or not two pfaffian systems are equivalent, but the algebraic operations involved in their application are in general too complicated to carry out, whereas the comparison of arithmetical invariants will often settle the question.

The Ω's are used (§8) to give a criterion for reducibility to a canonical system of a particular type, designated as *completely separable*, and to state necessary and sufficient conditions for the equivalence of such systems.

The reader is assumed to be familiar with the contents of Goursat's treatise.‡

2. **The fundamental invariant forms.** Consider the pfaffian system

$$S:\ \omega^1 = 0, \omega^2 = 0, \cdots, \omega^r = 0. \tag{2.1}$$

* Presented to the Society, December 27, 1932; received by the editors February 18, 1933. The results in this paper are taken from a doctoral dissertation in mathematics presented at Duke University, June, 1933.

† C. Riquier, *Les Systèmes d'Equations aux Dérivées Partielles*, Paris, 1910.

‡ E. Goursat, *Leçons sur le Problème de Pfaff*, Paris, 1922.

Let the symbol $\Omega^{i_1i_2\cdots i_k}$ be defined as follows:

$$\Omega^{i_1i_2\cdots i_k} = \omega^1\omega^2\cdots\omega^r\omega'^{i_1}\omega'^{i_2}\cdots\omega'^{i_k}, \tag{2.2}$$

where $i_1i_2\cdots i_k$ represents any combination of numbers selected from $1, 2, \cdots, r$ and ω' is the derived form of ω. The forms $\Omega^{i_1i_2\cdots i_k}$ are invariants of the system S. Moreover they are symmetric in every pair of superscripts.

Assuming the non-singular transformation

$$\bar{\omega}^i = a^i_\alpha\omega^\alpha,\ a = |a^i_j| \neq 0, \tag{2.3}$$

we have

$$\bar{\omega}'^i = a^i_\alpha\omega'^\alpha + da^i_\alpha\omega^\alpha,$$

where the repeated indices on the right indicate summation over the range $1, 2, \cdots, r$. This transformation induces on the Ω's the linear homogeneous transformation

$$\bar{\Omega}^{i_1i_2\ldots i_k} = a\,a^{i_1}_{\alpha_1}a^{i_2}_{\alpha_2}\cdots a^{i_k}_{\alpha_k}\Omega^{\alpha_1\alpha_2\ldots\alpha_k}. \tag{2.4}$$

Conversely, there exists a transformation (2.3) which induces a given transformation

$$\bar{\Omega}^{i_1i_2\ldots i_k} = B^{i_1i_2\ldots i_k}_{\alpha_1\alpha_2\ldots\alpha_k}\Omega^{\alpha_1\alpha_2\ldots\alpha_k} \tag{2.5}$$

on the Ω's provided the a's satisfy

$$k!B^{i_1i_2\ldots i_k}_{\alpha_1\alpha_2\ldots\alpha_k} = aP(a^{i_1}_{\alpha_1}a^{i_2}_{\alpha_2}\cdots a^{i_k}_{\alpha_k}), \tag{2.6}$$

where P indicates the summation of all terms obtained by permuting $\alpha_1\alpha_2\cdots\alpha_k$. The conditions for compatibility of (2.6) are

$$k!B^{i_1i_2\ldots i_k}_{\alpha_1\alpha_2\ldots\alpha_k} = P((B^{i_1i_1\ldots i_1}_{\alpha_1\alpha_1\ldots\alpha_1}B^{i_2i_2\ldots i_2}_{\alpha_2\alpha_2\ldots\alpha_2}\cdots B^{i_ki_k\ldots i_k}_{\alpha_k\alpha_k\ldots\alpha_k})^{1/k}). \tag{2.7}$$

If $k=1$, conditions (2.7) are identically satisfied.

3. Linear dependence of the Ω's. Expressing that the relations

$$C_{\alpha_1\alpha_2\ldots\alpha_k}\Omega^{\alpha_1\alpha_2\ldots\alpha_k} = 0 \tag{3.1}$$

are identically satisfied in the differentials gives a set of linear homogeneous equations on the C's, whose rank can be proved invariant under transformations (2.3). We shall employ Sylvester's term "nullity" as a name for the invariant p_k, which is the number of linearly independent solutions of (3.1).

In the case of the Ω's with a single index, by a transformation (2.3) an Ω

can be made zero corresponding to each relation (3.1). This gives the theory of the derived system,* which is an invariant system of the original.

Suppose it is possible to make the following p Ω's of order $k>1$ vanish:

$$\Omega^{11\cdots 1}=0,\ \Omega^{22\cdots 2}=0,\ \cdots,\ \Omega^{pp\cdots p}=0, \tag{3.2}$$

where p is the corresponding nullity. It can be proved by methods similar to those developed in greater detail in §8 that any transformation (2.3) which leaves (3.2) invariant permutes the first p equations of S among themselves. Those equations therefore form an invariant system of S.

4. **Further invariant systems associated with a pfaffian system.** All the Ω's of sufficiently high order for a given system are zero since the degree of $\Omega^{i_1 i_2\cdots i_k}$ finally exceeds the class of the system. Suppose every Ω with $\rho+1$ indices is identically zero, whereas some Ω with ρ indices does not vanish. Then 2ρ is an arithmetical invariant of the system. It is, in fact, easily identified with the invariant defined in a different manner by Engel† and called by him the rank of the system. For $q\leqq\rho$ let S_q be defined as the system composed of the equations formed by setting all common factors of $\Omega^{i_1 i_2\cdots i_q}$ equal to zero. The system S_q is always contained in S_{q+1}. We have then a sequence of invariant pfaffian systems all of whose arithmetical invariants are also invariants for S.

THEOREM 1. *A system S is of species one if and only if the corresponding S_1 is passive and does not coincide with S.*‡

THEOREM 2. *If S is of rank two, it can be put in a form satisfying*

$$\begin{gathered}\omega'^1\equiv 0,\ \cdots,\ \omega'^{r-3}\equiv 0,\ \omega'^{r-2}\equiv\phi^2\phi^3,\\ \omega'^{r-1}\equiv\phi^3\phi^1,\ \omega'^r\equiv\phi^1\phi^2,\ \mathrm{mod}\ \omega^1,\omega^2,\cdots,\omega^r,\end{gathered} \tag{4.1}$$

or

$$\begin{gathered}\omega'^1\equiv 0,\ \cdots,\ \omega'^{r'}\equiv 0,\ \omega'^{r'+1}\equiv\phi\psi^{r'+1},\ \cdots,\ \omega'^r\equiv\phi\psi^r,\\ \mathrm{mod}\ \omega^1,\omega^2,\cdots,\omega^r.\end{gathered} \tag{4.2}$$

In the first case, $S_1=S$; in the second, $S_1>S$. Conversely, if $S_1>S$, then S is of rank two and satisfies (4.2).

The proof of Theorem 2 follows. Since S is of rank two, every Ω of order two must vanish. The vanishing of Ω^{ij} when $i=j$ indicates that every ω'^i must vanish or be of rank two mod $\omega^1,\cdots,\omega^r$; thus the derived form of every ω

* Goursat, p. 294.

† F. Engel, Leipziger Berichte, vol. 52 (1890). This invariant is zero if and only if the system is passive.

‡ For the definition of species and for material which facilitates the proof of the above theorem, see J. M. Thomas, *Pfaffian systems of species one*, these Transactions, vol. 35 (1933).

not contained in the derived system is of rank two mod $\omega^1, \cdots, \omega^r$. Since this is true, the vanishing of Ω^{ij} when $i \neq j$ implies that every pair of non-vanishing derived forms possesses a common factor. This is possible in only two ways: S must satisfy (4.1) or (4.2). When (4.1) is satisfied, the derived system contains $r-3$ equations.

5. **Primitive systems.** If Σ has S for its derived system, Σ will be called a *primitive system* of S.

THEOREM 3. *Every pfaffian system has a primitive system.*

Let us assume that the derived system of

$$(5.1) \qquad S: \ \omega^1 = 0, \cdots, \omega^{r'} = 0, \cdots, \omega^r = 0$$

is

$$(5.2) \qquad S^1: \ \omega^1 = 0, \cdots, \omega^{r'} = 0.$$

The first r' derived forms of S then vanish by virtue of the system. When reduced by (5.1), the non-vanishing forms are quadratic in $\phi^1, \cdots, \phi^{n-r}$, which with the ω's constitute an independent set of forms. Consequently, they vanish by virtue of

$$(5.3) \qquad \omega^{r+1} = \phi^2 - \lambda^2\phi^1 = 0, \cdots, \omega^{n-1} = \phi^{n-r} - \lambda^{n-r}\phi^1 = 0.$$

If in addition the λ's form with the original variables an independent set, the derived form of no left member of this set of equations can vanish by virtue of S and (5.3). Hence the system $\Sigma = S + (5.3)$ is a primitive system of S.

THEOREM 4. *The minimum number of equations which adjoined to S yield a primitive system is an invariant of S.*

Let primitive systems for two equivalent systems S, $\bar{S}$ be

$$(5.4) \qquad \Sigma: \ \omega^1 = 0, \cdots, \omega^r = 0, \phi^1 = 0, \cdots, \phi^k = 0,$$

$$(5.5) \qquad T: \ \bar{\omega}^1 = 0, \cdots, \bar{\omega}^r = 0, \psi^1 = 0, \cdots, \psi^l = 0,$$

respectively, where k and l are least. The derived forms of S, being linear homogeneous combinations of those of $\bar{S}$, vanish whenever those of $\bar{S}$ do, and vice versa. This shows that $l \leqq k$ and $k \leqq l$, whence $k = l$.

Every passive system (species zero) of r equations is the derived system of a system of $r+1$ equations. When the adjunction of a single equation to a system of species one furnishes a primitive system is answered by the following:

THEOREM 5. *Every system of r equations whose species is one is the derived system of $2r-r'$ equations, but of no smaller number.*

A transformation (2.3) and a change of variables will put any system of species one in the form

$$\omega^1 = \cdots = \omega^{r'} = 0,\ dx^{r'+1} - A^{r'+1}dx^{r+1} = \cdots = dx^r - A^r dx^{r+1} = 0, \tag{5.6}$$

where the first r' equations constitute the derived system. A system has (5.6) in its derived system if and only if it implies the vanishing of the forms $dA^\alpha dx^{r+1}$, that is, if and only if it implies

$$dA^\alpha - \lambda^\alpha dx^{r+1} = 0. \tag{5.7}$$

If the forms dA^α were linearly dependent by virtue of (5.6), there would be more than r' equations in the derived system of (5.6). Therefore equations (5.7) form with (5.6) an independent set of $2r-r'$ equations, and no primitive system contains fewer equations. If x, A, λ constitute a set of independent variables, the system composed of (5.6) and (5.7) is a primitive system of (5.6).

THEOREM 6. *When the species exceeds one, the adjunction of a single equation gives a primitive system if and only if the class is $2r-r'+1$ and the rank two.*

If we put

$$\omega'^i \equiv G^i \mod \omega^1, \cdots, \omega^r \qquad (i = r'+1, \cdots, r),$$

the conditions of the theorem can be restated: the class of the set G^i is $r-r'+1$, and

$$G^i G^i = 0 \qquad (i = r'+1, \cdots r). \tag{5.8}$$

If the adjunction of

$$\phi = 0 \tag{5.9}$$

gives a primitive system,

$$G^i = \phi\psi^i.$$

Hence (5.8) are satisfied. If the forms ω, ϕ, ψ were not independent, the G's would be linearly dependent and the derived system would contain more than r' equations. Hence ω, ϕ, ψ are independent and the class of the system G^i is $r-r'+1$. Conversely, when (5.8) are satisfied, Theorem 2 shows that S can be displayed as (4.1) or (4.2). The class of (4.1) is $r+3$, however, whereas the expression $2r-r'+1$ reduces to $r+4$ when $r'=r-3$. Hence S is in the form (4.2), and the ϕ of those formulas furnishes a primitive system.

6. **A theorem on matrices.** A square matrix is *monomial* if it contains one and only one non-zero element on each row and on each column. The totality of monomial matrices of a given order has the group property under multiplication. It will be called the *monomial group.*

LEMMA. *If a non-singular square matrix is multiplied by a properly chosen monomial matrix, every element in the main diagonal of the resulting matrix is different from zero.*

Let the given matrix be M. Consider any non-vanishing term of the expansion of the determinant $|M|$. In the matrix M replace each element of this term by unity and every other element of the matrix by zero; call the transpose of the resulting matrix N. Then MN has the chosen non-zero elements on its main diagonal.

THEOREM 7. *If the elements a_i^j of a non-singular square matrix M satisfy the conditions*

$$P(a_1^{i_1}a_2^{i_2}\cdots a_r^{i_r}) = 0, \tag{6.1}$$

where the $i_1, \cdots, i_r$ are any set of values from the range $1, 2, \cdots, r$ such that at least two of them are equal and P indicates the summation of all the terms obtained by permuting the subscripts, then M is monomial.

Consider the product MN, where N is constructed as in the preceding lemma. This matrix MN also satisfies conditions (6.1) because multiplication on the right by N simply permutes the columns, thus permuting the subscripts in (6.1). Hence MN is a matrix satisfying (6.1) and also

$$a_1^1a_2^2\cdots a_r^r \neq 0. \tag{6.2}$$

If MN is monomial, then M is too. Therefore it suffices to show that any matrix satisfying conditions (6.1) and (6.2) is monomial.

The theorem holds for $r=2$ because conditions (6.1) and (6.2) are

$$a_1^1a_2^1 = 0, \quad a_1^2a_2^2 = 0; \quad a_1^1a_2^2 \neq 0.$$

Assume the theorem true for every matrix of order less than r and suppose (6.1), (6.2) satisfied by a matrix of order r. From (6.1) we have $a_1^1a_2^1\cdots a_r^1=0$, whence $a_2^1a_3^1\cdots a_r^1=0$. To prove

$$a_2^1 = a_3^1 = \cdots = a_r^1 = 0 \tag{6.3}$$

we employ induction. Suppose $k-1$ elements on the first row are zero. By interchanging, if necessary, certain rows and the corresponding columns, we make those elements $a_2^1, a_3^1, \cdots, a_k^1$ and preserve the condition (6.2). From (6.1),

$$P(a_1^1a_2^{i_2}\cdots a_k^{i_k}a_{k+1}^1\cdots a_r^1) = 0,$$

where $i_2, \cdots, i_k$ are chosen from the range $2, \cdots, k$ in every possible way. Since $a_2^1 = a_3^1 = \cdots = a_k^1 = 0$ and $a_1^1 \neq 0$, this reduces to

$$P(\overset{i_2}{a_2} \cdots \overset{i_k}{a_k})\overset{1}{a}_{k+1} \cdots \overset{1}{a_r} = 0.$$

We wish to show that

$$\overset{1}{a}_{k+1} \cdots \overset{1}{a_r} = 0 \tag{6.4}$$

holds. Assuming the contrary, we have

$$P(\overset{i_2}{a_2} \cdots \overset{i_k}{a_k}) = 0. \tag{6.5}$$

But the conditions of the theorem are then satisfied for the range $2, \cdots, k$ and by assumption they imply

$$a_i^j = 0 \qquad (i, j = 2, 3, \cdots, k; i \neq j). \tag{6.6}$$

The condition $P(a_2^2 \cdots a_k^k) = 0$, where the upper indices are permuted, is also implied by (6.5). Because of (6.6) it reduces to $a_2^2 \cdots a_k^k = 0$. This contradiction forces us to conclude that (6.4) is true. Hence we may assume $\overset{1}{a}_{k+1} = 0$, and by induction we reach the result that all elements except a_1^1 in the first row are zero. Since any row can be made the first by a transformation that preserves (6.2), the same argument can subsequently be applied to each of the other rows to show that

$$a_i^j = 0 \qquad (i, j = 1, \cdots, r; i \neq j),$$

and the theorem therefore is true for the matrix of order r.

7. **Partition of pfaffian systems.** Suppose that for a pfaffian system certain relations

$$F_1 = 0, F_2 \neq 0, \cdots \tag{7.1}$$

are satisfied. Let $\mathcal{G}$ be the subgroup of transformations (2.3) leaving (7.1) invariant. If $\mathcal{G}$ is intransitive or imprimitive,* the transformation (2.3) which displays its ultimate sets of intransitivity or imprimitivity will exhibit S as the sum† of a number of pfaffian systems

$$S = S_1 + S_2 + \cdots + S_q.$$

This will be called a *partition* of S.

If $\mathcal{G}$ is intransitive, each S_i is an invariant system of S. If $\mathcal{G}$ is imprimitive,

* These terms are employed in the sense defined by H. F. Blichfeldt, *Finite Collineation Groups*, Chicago, 1917, p. 17 and p. 76.

† We apply the symbol + only to aggregates having no elements in common.

the systems S_i all contain the same number of equations and are permuted when S is subjected to the general transformation (2.3).*

If each of the systems S_i contains only one equation, the partition will be called *complete*; this occurs only when the group $\mathcal{G}$ is monomial.

THEOREM 8. *The conditions*

$$\Omega^{12\cdots r} \neq 0, \quad \Omega^{i_1 i_2 \cdots i_r} = 0 \quad (i_1 i_2 \cdots i_r \neq 12 \cdots r) \tag{7.2}$$

define a complete partition of the pfaffian system.

By (2.4) the conditions that the second of (7.2) be preserved under (2.3) are

$$a_{\alpha_1}^{i_1} a_{\alpha_2}^{i_2} \cdots a_{\alpha_r}^{i_r} \Omega^{\alpha_1 \alpha_2 \cdots \alpha_r} = 0 \quad (i_1 i_2 \cdots i_r \neq 12 \cdots r)$$

and reduce precisely to (6.1). The result follows from Theorem 7.

A system admitting partition is

$$dx^1 + x^2 dx^3 = 0, \quad dx^4 + x^7 dx^6 = 0, \quad dx^5 + x^8 dx^6 = 0. \tag{7.3}$$

For it the defining relations (7.1) are

$$\Omega^{11} = \Omega^{22} = \Omega^{33} = \Omega^{23} = 0, \quad p_2 = 4.$$

8. **Separable systems.** A system will be called *separable* if it is equivalent to a system S expressed in terms of a set of independent variables X such that

$$S = S_1 + S_2, \quad X = X_1 + X_2,$$

where S_1 is expressed in terms of the variables X_1, S_2 in terms of X_2, and neither S_1 nor S_2 is vacuous.

The system is *completely separable* if it is equivalent to

$$S = S_1 + S_2 + \cdots + S_r,$$

expressed in terms of variables

$$X = X_1 + X_2 + \cdots + X_r,$$

each S_i containing a single equation and being expressed in terms of the variables in the corresponding X_i. The following is readily proved:

THEOREM 9. *A completely separable system can be written in a canonical form each equation of which is in the canonical form for a single equation.*

Generalization of a known method† proves

* Invariant systems like those of §3 arise when $\mathcal{G}$ is merely reducible.

† Cf. Goursat, p. 308, where the method is employed in reducing a system of two equations in four variables to canonical form.

THEOREM 10. *A pfaffian system of r equations having $r-1$ independent integrals is completely separable.*

We shall now derive a necessary and sufficient condition that a system be completely separable. Suppose a completely separable system written in canonical form. Let $\omega'^i \equiv G^i$, mod $\omega^1, \cdots, \omega^r$. The number of equations in the derived system is the number of independent solutions of

$$(8.1) \qquad \lambda_\alpha G^\alpha = 0.$$

Since $G^1, \cdots, G^r$ have no differentials in common,

$$(8.2) \qquad \lambda_1 G^1 = 0, \cdots, \lambda_r G^r = 0.$$

Suppose exactly q of the G's are zero. They can be made $G^1, \cdots, G^q$, and the first q equations of S are

$$(8.3) \qquad dx^1 = 0, \cdots, dx^q = 0.$$

The remaining λ's are zero and the number of independent solutions of (8.1) is q. Hence (8.3) is the derived system S^1 of S. The derived system of a completely separable system is therefore passive. We have

$$(8.4) \qquad \Omega^1 = 0, \cdots, \Omega^q = 0,$$

whereas the remaining Ω's of the first order are linearly independent.

Let the class of ω^i be $2m_i+1$. It is easily verified that every Ω of order $m_1+m_2+\cdots+m_r$ is zero except

$$(8.5) \qquad \Omega^{(q+1)(q+1)\ldots(q+1)\ldots rr\ldots r} \neq 0$$

where the superscript i occurs m_i times. We denote the conditions that the Ω's other than (8.5) vanish by

$$(8.6) \qquad \Omega_{m_1+m_2+\cdots+m_r} = 0.$$

Consider a transformation (2.3) which leaves the two sets of conditions above invariant. The preservation of (8.4) gives

$$(8.7) \qquad a_j^i = 0 \qquad (i = 1, 2, \cdots, q;\ j = q+1, \cdots, r).$$

The preservation of the other conditions (8.6) gives

$$(8.8) \qquad P(a_{q+1}^{\alpha_1} a_{q+1}^{\alpha_2} \cdots a_{q+1}^{\alpha_{m_{q+1}}} a_{q+2}^{\beta_1} a_{q+2}^{\beta_2} \cdots a_{q+2}^{\beta_{m_{q+2}}} \cdots a_r^{\gamma_1} a_r^{\gamma_2} \cdots a_r^{\gamma_{m_r}}) = 0,$$

where $\alpha, \beta, \cdots, \gamma$ have any values from the range $q+1, \cdots, r$ except the values occurring in (8.5). In particular, when all the α's are equal, all the β's are equal, etc., we have

(8.9) $$P\{(a_{q+1}^{\alpha})^{m_{q+1}}(a_{q+2}^{\beta})^{m_{q+2}}\cdots(a_r^{\gamma})^{m_r}\} = 0,$$

where $\alpha, \beta, \cdots, \gamma$ is any set from $q+1, \cdots, r$ which is not a permutation of $q+1, \cdots, r$ and the m's denote powers.

The substitution

(8.10) $$(a_{q+1}^{\alpha})^{m_{q+1}} = b_{q+1}^{\alpha}, (a_{q+2}^{\beta})^{m_{q+2}} = b_{q+2}^{\beta}, \cdots, (a_r^{\gamma})^{m_r} = b_r^{\gamma}$$

gives

$$P(b_{q+1}^{\alpha}b_{q+2}^{\beta}\cdots b_r^{\gamma}) = 0.$$

Hence the theorem of §6 can be applied to show that the matrix

$$\|b_i^j\| \qquad (i, j = q+1, \cdots, r)$$

and consequently the matrix

$$\|a_i^j\| \qquad (i, j = q+1, \cdots, r),$$

is monomial.

If we write

(8.11) $$S = S^1 + S_{q+1} + S_{q+2} + \cdots + S_r,$$

the system S^1 is left invariant by any transformation (2.3) preserving (8.4), (8.5), (8.6); and the systems

(8.12) $$S^1 + S_{q+1}, S^1 + S_{q+2}, \cdots, S^1 + S_r,$$

each of which contains $q+1$ equations, are permuted among themselves by such a transformation. The class values for the canonical form of systems (8.12) are

(8.13) $$q + 2m_{q+1} + 1, q + 2m_{q+2} + 1, \cdots, q + 2m_r + 1,$$

and consequently the class values must be these whenever conditions (8.4), (8.5), (8.6) are satisfied. Thus we have an additional necessary condition on a completely separable system.

The conditions given above are also sufficient. By Theorem 10 systems (8.12) can be written in canonical form. From the result already established concerning the derived system of a completely separable system in canonical form, the first q equations in each of the systems (8.12) must be equations (8.2). Since the class values are the set (8.13), the system is expressed in terms of $x^1, \cdots, x^q$ and $2(m_{q+1} + \cdots + m_r) + r - q$ other variables. Since the total number of these variables is the same as the degree of the Ω in (8.5), that Ω is the product of their differentials, and its non-vanishing declares the set of variables independent. Thus we have

THEOREM 11. *A pfaffian system is completely separable into r equations of class* $1, \cdots, 1, 2m_{q+1}+1, \cdots, 2m_r+1$, *where no m is zero, if and only if the following conditions are satisfied. It must be possible to determine a transformation* (2.3) *which realizes* (8.4), (8.5), (8.6). *When such a transformation* (2.3) *has been found and applied to the system, the numbers* (8.13) *must be the class values of* (8.12).

The determination of the transformation (2.3) involves finding a particular solution of a system of homogeneous algebraic equations in the a's whose degree is $m_{q+1}+ \cdots +m_r$. It is important to note that the second condition in Theorem 11 is either satisfied for all solutions of this system of algebraic equations or for none.

A completely separable system having no integrals admits a complete partition. That the converse is not true is evident from (7.3).

DUKE UNIVERSITY,
DURHAM, N. C.

THE COMPLETE EXISTENTIAL THEORY OF THE WHITEHEAD-HUNTINGTON SET OF POSTULATES FOR THE ALGEBRA OF LOGIC*

BY

A. H. DIAMOND

1. **Introduction.** Consider any set of postulates, say, for the sake of concreteness, P_1, P_2, P_3. Any interpretation of the undefined ideas of P_1, P_2, P_3 constitutes a concrete system S which either satisfies or does not satisfy some or all of the postulates. The system S will then have with respect to P_1, P_2, P_3 one of the 2^3 *characters* $(\pm\pm\pm)$, where a "+" sign in the ith place denotes that postulate P_i is satisfied and a "−" sign that P_i is not satisfied. The *complete existential theory* of the postulate-set P_1, P_2, P_3 consists in determining for every one of the 2^3 characters $(\pm\pm\pm)$ whether or not there exists a concrete system S corresponding to that character.† The object of this paper is to establish the complete existential theory of the Whitehead-Huntington set of ten postulates for the algebra of logic, expressed in terms of logical addition and logical multiplication.‡ This is, perhaps, the most "natural" and most elegant set of postulates for the Boole-Schröder algebra of logic.

The complete existential theory of a set of postulates includes the solution of the problem of determining whether or not the postulates of the set (or of any of its sub-sets) are *completely independent*, i.e., whether or not any of the postulates of the set (or sub-set) or their denials can be derived from any of the other postulates or their denials. Thus the present discussion will show whether or not the postulates of the Whitehead-Huntington set (or any of its sub-sets) are completely independent.

The present theory has the following distinctive characteristics. (1) The number of postulates involved is 10, and so requires, for the establishment of the complete existential theory, 2^{10} propositions of existence and non-existence. This number is far greater than the number of propositions which constitute any complete existential theory hitherto published.§ (2) The con-

* Presented to the Society, March 18, 1933; received by the editors March 18, 1933.

† Professor E. H. Moore first proposed the problem of the complete existential theory of a set of postulates. See his *Introduction to a Form of General Analysis*, New Haven Colloquium, Yale University Press, p. 82.

‡ See E. V. Huntington, these Transactions, vol. 5 (1904), pp. 288–309.

§ The largest number hitherto published is $2^6=64$. See B. A. Bernstein, *The complete existential theory of Hurwitz's postulates for abelian groups and fields*, Bulletin of the American Mathematical

crete systems employed are all in the *modular* form devised by Professor B. A. Bernstein.* This form permits the listing of the large number of systems involved with a conciseness and a simplicity not existing in the proof-systems employed before the development of the modular theory in question. (3) The number of systems constituting the propositions of existence is 325. This number is far larger than any previous number of proof-systems employed in connection with a set of postulates. Further, the systems relate to postulates expressing laws found in many other mathematical theories. These systems thus provide a large store of possible proof-systems for many important postulate-sets. (4) The systems employed are all algebras of not more than three elements. This adds to the value of the systems as a source of possible proof-systems to be used in other postulate-sets.

I begin with the listing of the Whitehead-Huntington postulates.

2. **The Whitehead-Huntington postulates.** The Whitehead-Huntington set of postulates for the algebra of logic leave undefined a *class* K, and two *binary operations* $+$, $\times$, and are the ten propositions following.†

Ia. $a+b$ is in K whenever a and b are in K.

Ib. ab is in K whenever a and b are in K.

IIa. There is an element Z such that $a+Z=a$ for every element a.

IIb. There is an element U such that $aU=a$ for every element a.

IIIa. $a+b=b+a$ whenever a, b, $a+b$, and $b+a$ are in K.

IIIb. $ab=ba$ whenever a, b, ab, and ba are in K.

IVa. $a+bc=(a+b)(a+c)$ whenever a, b, c, $a+b$, $a+c$, bc, $a+bc$, and $(a+b)(a+c)$ are in K.

IVb. $a(b+c)=ab+ac$ whenever a, b, c, ab, ac, $b+c$, $a(b+c)$, and $ab+ac$ are in K.

V. If the elements Z and U in postulates IIa and IIb exist and are unique, then for every element a there is an element a' such that $a+a'=U$ and $aa'=Z$.

VI. There are at least two elements, x and y, in K such that $x\neq y$.

Society, vol. 28 (1922), p. 397. $2^5=32$ propositions occur in two other papers. See Paul Henle, *The independence of the postulates of logic*, Bulletin of the American Mathematical Society, vol. 38 (1932), p. 409. See also J. S. Taylor, *Sheffer's set of five postulates for Boolean algebras in terms of the operation "rejection" made completely independent*, Bulletin of the American Mathematical Society, vol. 26 (1920), p. 449.

* See B. A. Bernstein, *Modular representations of finite algebras*, Proceedings of the International Mathematical Congress, Toronto, 1924, p. 207. See also B. A. Bernstein and Nemo Debely, *A practical method for the modular representation of finite operations and relations*, Bulletin of the American Mathematical Society, vol. 38 (1932), p. 110.

† The original wording of the postulates is retained except that K replaces Huntington's "class", the circles around the operations $+$ and $\times$ are omitted, and the original $\wedge$, $\vee$, and $\bar{a}$ are replaced by Z, U and a' respectively.

If these postulates be denoted by $a_1, a_2, \cdots, a_{10}$ and their denials by $a_1', a_2', \cdots, a_{10}'$, then the propositions $a_1, a_2, \cdots, a_{10}, a_1', a_2', \cdots, a_{10}'$ divide the universe of discourse of these propositions into $2^{10}=1024$ compartments represented by the logical products $a_1a_2 \cdots a_{10}, a_1a_2 \cdots a_{10}', \cdots, a_1'a_2' \cdots a_{10}'$.* Only 325 of these compartments are actually represented in the universe; the remaining 699 are empty because of the relations of implication subsisting among the propositions $a_1, a_2, \cdots, a_{10}, a_1', a_2', \cdots, a_{10}'$.

3. **Propositions of non-existence.** Of the 2^{10} propositions constituting the complete existential theory of our postulates, $2^{10}-325=699$ are propositions of non-existence. These non-existence propositions, together with reasons establishing them, are given by propositions A–F following.

A. *There exist no systems for the characters* $(\pm\pm\pm\pm\pm\pm\pm\pm\pm-)$ *except* $(\pm\pm\pm\pm+++++-)$.

For, if postulate VI is denied by a system S, then K has just one element or none. If K has no elements, S satisfies postulates IIIa, IIIb, IVa, IVb, and V vacuously. If K has just one element, there are four cases to consider. (1) If S satisfies both postulates Ia and Ib, then S satisfies postulates IIIa, IIIb, IVa, IVb, and V non-vacuously. (2) If S denies both postulates Ia and Ib, then S satisfies postulates IIIa, IIIb, IVa, IVb, and V vacuously. (3) If S satisfies postulate Ia and denies postulate Ib, then S satisfies postulate IIIa non-vacuously and postulates IIIb, IVa, IVb, and V vacuously. (4) If S denies postulate Ia and satisfies postulate Ib, then S satisfies IIIb non-vacuously and postulates IIIa, IVa, IVb, and V vacuously.

Proposition A accounts for 496 characters.

B. *There exist no systems for the characters* $(-\pm+\pm\pm\pm\pm\pm\pm-)$.

For, if postulate VI is denied, postulate IIa will be satisfied if and only if K has just one element and S satisfies postulate Ia.

Proposition B accounts for 4 characters not already accounted for by proposition A, namely $(-\pm+\pm+++++-)$.

C. *There exist no systems for the characters* $(\pm-\pm+\pm\pm\pm\pm\pm-)$.

For, if postulate VI is denied, postulate IIb will be satisfied if and only if K has just one element and S satisfies postulate Ib.

Proposition C accounts for 3 characters not already accounted for by propositions A and B, $(\pm--++++++-)$ and $(+-+++++++-)$.

D. *There exist no systems for the characters* $(+++-\pm\pm\pm\pm\pm-)$ *and* $(++-+\pm\pm\pm\pm\pm-)$.

* See E. V. Huntington, these Transactions, vol. 26 (1924), p. 277.

For, if postulate VI is denied, there are two ways in which postulates Ia and Ib may both be satisfied. (1) If K has no elements, postulates Ia and Ib are satisfied vacuously. (2) If K has just one element, postulates Ia and Ib may both be satisfied non-vacuously. In case (1) both postulates IIa and IIb are denied. In case (2) both postulates IIa and IIb are satisfied.

Proposition D accounts for 2 characters not already accounted for by propositions A, B, and C, namely $(+++-+++++-)$ and $(++-++ ++++-)$.

E. *There exist no systems for the characters* $(+---\pm\pm\pm\pm\pm-)$ *and* $(-+--\pm\pm\pm\pm\pm-)$.

For, if postulate VI is denied, there are two ways in which postulates IIa and IIb may both be denied. (1) If K has no elements, postulates IIa and IIb are denied non-vacuously and postulates Ia and Ib are satisfied vacuously. (2) If K has just one element, postulates IIa and IIb are both denied if and only if postulates Ia and Ib are both denied.

Proposition E accounts for 2 characters not already accounted for by propositions A, B, C, and D, namely $(+---+++++-)$ and $(-+-- +++++-)$.

F. *There exist no systems for the characters* $(\pm\pm+-\pm\pm\pm\pm-\pm)$, $(\pm\pm-+\pm\pm\pm\pm-\pm)$, *and* $(\pm\pm--\pm\pm\pm\pm-\pm)$.

For, if postulate IIa or IIb or both are denied, then postulate V is satisfied vacuously.

Proposition F accounts for 192 characters not already accounted for by propositions A, B, C, D, and E, namely $(\pm\pm+-\pm\pm\pm\pm-+)$, $(\pm\pm -+\pm\pm\pm\pm-+)$ and $(\pm\pm--\pm\pm\pm\pm-+)$.

4. **Propositions of existence.** The 325 propositions of existence for our postulates are given by the tables A and B below. In these tables all the systems are *arithmetic* systems, the elements being the numbers 0, 1, and 2. The notations used are those employed by Professor Bernstein in the second of the papers cited in the first footnote on page 941, except that, for the sake of saving space, certain abbreviations are resorted to. Let $f(a, b)$ be any polynomial expression in a and b, where a and b are any of the numbers $0, 1, \cdots, p-1$. Then $(f(a, b))_p$ will denote the *least positive residue modulo* p obtained from $f(a, b)$ by rejecting multiples of p. The operations $+$ and $\times$ are to be interpreted as the operations of ordinary arithmetic when they occur in the modular expressions, otherwise they are to be interpreted as logical addition and logical multiplication. Thus for $a=1$, $b=1$, we have $(a^2+ab+b+2)_3=2$. $[a, b; m, n]_p$ will denote a function $(f(a, b))_p$ such that $(f(a, b))_p=0$ or 1 according as the equalities $a=m$, $b=n$ do or do not both hold. The results

obtained by Professor Bernstein enable us to write down the expression for $[a, b; m, n]_p$. Indeed, $[a, b; m, n]_p = (1 - \{1-(a-m)^{p-1}\}\{1-(b-n)^{p-1}\})_p$. Thus $[a, b; 1, 2]_3 = (1-\{1-(a-1)^2\}\{1-(b-2)^2\})_3 = (2a^2b^2+a^2b+2ab^2 +ab+1)_3$.

For the sake of simplicity, the 325 systems will be divided into two groups, A and B, according as K does not have or does have more than one element. Table A gives systems for which K does not have more than one element; table B gives systems for which K does have more than one element. To save space in table B, instead of writing out the elements of K as in table A, the subscript p in $(f(a, b))_p$ will indicate that the elements of K are $0, 1, \cdots, p-1$. In both tables, the characters will be written without parentheses.

The duality of the postulates with respect to the operations $+$ and $\times$ makes it possible to reduce the number of systems in table B from 320 to 172. Every system S has a dual system obtained from S by interchanging the definitions of $+$ and $\times$. The character corresponding to the dual of a system S is determined as follows. Consider the character C corresponding to S. C consists of "+" and "−" signs arranged in a certain order. The first 8 signs occur in pairs corresponding to dual postulates. Interchange in C the signs in each pair. The character thus obtained is the character corresponding to the dual of S.

TABLE A

No.	Character	K	$a+b$	ab
(i)	− − − − + + + + + + −	0	0/0	0/0
(ii)	− + − + + + + + + + −	0	0/0	0
(iii)	+ − + − + + + + + + −	0	0	0/0
(iv)	+ + − − + + + + + + −	Null	——	——
(v)	+ + + + + + + + + + −	0	0	0

TABLE B

No.	Character										$a+b$	ab
	Ia	Ib	IIa	IIb	IIIa	IIIb	IVa	IVb	V	VI		
1	−	−	−	−	−	−	−	−	+	+	$(a+1)_2+0/[a, b; 1, 1]_2$	$(ab+b)_2+0/[a, b; 0, 0]_2$
2	−	−	−	−	−	−	−	+	+	+	$a+b$ of system 1	$(ab+a)_2+0/[a, b; 0, 0]_2$
3	−	−	−	−	−	−	+	+	+	+	$(b)_2+0/[a, b; 1, 1]_2$	$a+b$ of this system
4	−	−	−	−	−	+	−	−	+	+	• $(b+1)_2+0/[a, b; 1, 1]_2$	$0/[a, b; 1, 1]_2$
5	−	−	−	−	−	+	−	+	+	+	$a+b$ of system 1	$(ab+1)_2+0/[a, b; 0, 0]_2$
6	−	−	−	−	−	+	+	−	+	+	" " " "	ab of system 4
7	−	−	−	−	−	+	+	+	+	+	" " " 3	" " " "
8	−	−	−	−	+	+	−	−	+	+	$0/[a, b; 1, 0]_2$	$1+0/[a, b; 1, 1]_2$

TABLE B (*Continued*)

No.	Character										$a+b$	ab
	Ia	Ib	IIa	IIb	IIIa	IIIb	IVa	IVb	V	VI		
9	−	−	−	−	+	+	−	+	+	+	ab of system 4	ab of system 8
10	−	−	−	−	+	+	+	+	+	+	$(0/0)_2$	$(0/0)_2$
11	−	−	−	+	−	−	−	−	+	+	$(a+1)_3+0/[a, b; 2,2]_3$	$(ab+a)_3+0/[a, b; 2,2]_3$
12	−	−	−	+	−	−	−	+	+	+	$(2a^2+b+1)_3+0/[a, b; 2,2]_3$	$(2a^2b+a+b)_3+0/[a, b; 2,2]_3$
13	−	−	−	+	−	−	+	−	+	+	$a+b$ of system 1	$(a)_2+0/[a, b; 1, 1]_2$
14	−	−	−	+	−	−	+	+	+	+	" " " 3	ab of system 13
15	−	−	−	+	−	+	−	−	+	+	$(b+1)_3+0/[a, b; 0, 0]_3$	$(ab)_3+0/[a, b; 0, 0]_3$
16	−	−	−	+	−	+	−	+	+	+	$a+b$ of system 1	$(a+b)_2+0/[a, b; 0, 1]_2$
17	−	−	−	+	−	+	+	−	+	+	" " " "	$(a)_2+0/[a, b; 1, 0]_2$
18	−	−	−	+	−	+	+	+	+	+	" " " "	$(a+b)_2+0/[a, b; 1, 1]_2$
19	−	−	−	+	+	−	−	−	+	+	$1+0/[a, b; 2, 2]_3$	ab of system 11
20	−	−	−	+	+	−	−	+	+	+	$(2a+2b)_3+0/[a, b; 2, 2]_3$	" " " 12
21	−	−	−	+	+	−	+	−	+	+	ab of system 8	" " " 13
22	−	−	−	+	+	−	+	+	+	+	" " " 4	" " " "
23	−	−	−	+	+	+	−	−	+	+	$(a+b)_2+0/[a, b; 0, 0]_2$	" " " 16
24	−	−	−	+	+	+	−	+	+	+	$(b+1)_2+0/[a, b; 1, 0]_2$	" " " 18
25	−	−	−	+	+	+	+	−	+	+	ab of system 8	$(ab)_2+0/[a, b; 1, 0]_2$
26	−	−	−	+	+	+	+	+	+	+	" " " 4	$(ab)_2+0/[a, b; 0, 0]_2$
27	−	−	+	+	−	−	−	−	−	+	" " " 11	ab of system 11
28	−	−	+	+	−	−	−	−	+	+	$(ab+a)_3+0/[a, b; 1, 1]_3$	$a+b$ of this system
29	−	−	+	+	−	−	−	+	−	+	$(2a^2b^2+ab+a)_3+0/[a, b; 2,2]_3$	ab of system 12
30	−	−	+	+	−	−	−	+	+	+	$(a)_3+0/[a, b; 2, 2]_3$	" " " 11
31	−	−	+	+	−	−	+	+	−	+	ab of system 13	" " " 13
32	−	−	+	+	−	−	+	+	+	+	$a+b$ " " 30	$a+b$ " " 30
33	−	−	+	+	−	+	−	−	−	+	$(ab+a)_3+0/[a, b; 1, 2]_3$	$(ab)_3+0/[a, b; 1, 2]_3$
34	−	−	+	+	−	+	−	−	+	+	$(ab^2)_3+0/[a, b; 2, 0]_3$	$(a+b)_3+0/[a, b; 0, 2]_3$
35	−	−	+	+	−	+	−	+	−	+	$(a)_2+0/[a, b; 0, 0]_2$	ab of system 16
36	−	−	+	+	−	+	−	+	+	+	$a+b$ of system 30	$(a+b)_3+0/[a, b; 2, 2]_3$
37	−	−	+	+	−	+	+	−	−	+	$(2a^2b+a+b)_3+0/[a,b; 1,1]_3$	$(ab+2a+2b+2)_3+0/[a, b; 1, 1]_3$
38	−	−	+	+	−	+	+	−	+	+	$(ab^2)_3+0/[a, b; 0, 0]_3$	$(2a^2b+2ab^2+a+b)_3+0/[a,b; 2,2]_3$
39	−	−	+	+	−	+	+	+	−	+	ab of system 13	$(a)_2+0/[a, b; 0, 1]_2$
40	−	−	+	+	−	+	+	+	+	+	$a+b$ " " 30	$(2a^2b+a+b)_3+0/[a, b; 1, 2]_3$
41	−	−	+	+	+	+	−	−	−	+	ab " " 16	ab of system 18
42	−	−	+	+	+	+	−	−	+	+	" " " "	" " " 16
43	−	−	+	+	+	+	−	+	−	+	" " " 39	" " " "
44	−	−	+	+	+	+	−	+	+	+	$(a^2b+ab^2+a+b)_3+\sum_{i=1}^{2}0/[a, b; 0, i]_3$	$(a^2b+ab^2+a+b)_3+0/[a, b; 1, 1]_3$
45	−	−	+	+	+	+	+	+	−	+	ab of system 39	ab of system 39
46	−	−	+	+	+	+	+	+	+	+	" " " 18	" " " 26
47	−	+	−	−	−	−	−	−	+	+	$a+b$ " " 1	$(ab+b)_2$
48	−	+	−	−	−	−	−	+	+	+	" " " "	$(b+1)_2$
49	−	+	−	−	−	−	+	−	+	+	ab " " 2	$(a+1)_2$
50	−	+	−	−	−	−	+	+	+	+	$a+b$ " " 3	$(b)_2$
51	−	+	−	−	−	+	−	−	+	+	" " " 4	$(0)_2$
52	−	+	−	−	−	+	−	+	+	+	" " " 1	$(ab+1)_2$
53	−	+	−	−	−	+	+	−	+	+	$(2a^2)_3+0/[a, b; 0, 0]_3$	$(a^2+b^2)_2$

TABLE B (*Continued*)

No.	Character Ia	Ib	IIa	IIb	IIIa	IIIb	IVa	IVb	V	VI	$a+b$	ab
54	−	+	−	−	−	+	+	+	+	+	$a+b$ of system 3	ab of system 51
55	−	+	−	−	+	−	−	−	+	+	ab " " 4	" " " 48
56	−	+	−	−	+	−	−	+	+	+	" " " "	" " " 49
57	−	+	−	−	+	−	+	−	+	+	" " " 5	" " " "
58	−	+	−	−	+	−	+	+	+	+	" " " 4	" " " 50
59	−	+	−	−	+	+	−	−	+	+	$a+b$ " " 8	$(1)_2$
60	−	+	−	−	+	+	−	+	+	+	ab " " 4	ab of system 59
61	−	+	−	−	+	+	+	−	+	+	$(a+b+1)_2+0/[a, b; 0, 1]_2$	$(ab+a+b+1)_2$
62	−	+	−	−	+	+	+	+	+	+	ab of system 10	ab of system 51
63	−	+	−	+	−	−	−	−	+	+	$(b+1)_2+0/[a, b; 0, 0]_2$	$(ab+b+1)_2$
64	−	+	−	+	−	−	−	+	+	+	$a+b$ of system 63	$(ab+a)_2$
65	−	+	−	+	−	−	+	−	+	+	" " " 1	$(a)_2$
66	−	+	−	+	−	−	+	+	+	+	" " " 3	ab of system 65
67	−	+	−	+	−	+	−	−	+	+	" " " 63	$(ab+a+b)_2$
68	−	+	−	+	−	+	−	+	+	+	" " " 1	$(a+b)_2$
69	−	+	−	+	−	+	+	−	+	+	" " " "	$(ab)_2$
70	−	+	−	+	−	+	+	+	+	+	" " " 3	ab of system 68
71	−	+	−	+	+	−	−	−	+	+	ab of system 8	" " " 64
72	−	+	−	+	+	−	−	+	+	+	$a+b$ " " 20	$(2a^2b+a+b)_3$
73	−	+	−	+	+	−	+	−	+	+	ab " " 8	ab of system 65
74	−	+	−	+	+	−	+	+	+	+	" " " 4	" " " "
75	−	+	−	+	+	+	−	−	+	+	$a+b$ " " 23	" " " 68
76	−	+	−	+	+	+	−	+	+	+	" " " 24	" " " "
77	−	+	−	+	+	+	+	−	+	+	ab " " 8	" " " 69
78	−	+	−	+	+	+	+	+	+	+	" " " 4	" " " "
79	−	+	+	−	−	−	−	−	+	+	" " " 11	$(a+1)_3$
80	−	+	+	−	−	−	−	+	+	+	" " " 13	ab of system 49
81	−	+	+	−	−	−	+	−	+	+	" " " 12	$(2a^2+b+1)_3$
82	−	+	+	−	−	−	+	+	+	+	" " " 13	ab of system 50
83	−	+	+	−	−	+	−	−	+	+	" " " 11	$(1)_3$
84	−	+	+	−	−	+	−	+	+	+	" " " 13	ab of system 59
85	−	+	+	−	−	+	+	−	+	+	" " " 12	$(2a+2b)_3$
86	−	+	+	−	−	+	+	+	+	+	$(a)_3+0/[a, b; 0, 0]_3$	ab of system 85
87	−	+	+	−	+	−	−	−	+	+	$(a+b+1)_2+0/[a, b; 1, 0]_2$	" " " 47
88	−	+	+	−	+	−	−	+	+	+	ab of system 17	" " " 49
89	−	+	+	−	+	−	+	−	+	+	" " " 16	" " " "
90	−	+	+	−	+	−	+	+	+	+	" " " 18	" " " "
91	−	+	+	−	+	+	−	−	+	+	" " " 16	" " " 52
92	−	+	+	−	+	+	−	+	+	+	" " " 25	" " " 59
93	−	+	+	−	+	+	+	−	+	+	" " " 18	" " " 61
94	−	+	+	−	+	+	+	+	+	+	" " " 26	" " " 51
95	−	+	+	+	−	−	−	−	−	+	" " " 11	$(ab+a)_3$
96	−	+	+	+	−	−	−	−	+	+	$a+b$ " " 28	ab of system 95
97	−	+	+	+	−	−	−	+	−	+	ab " " 13	" " " 64
98	−	+	+	+	−	−	−	+	+	+	$a+b$ " " 30	" " " 95
99	−	+	+	+	−	−	+	−	−	+	ab " " 12	$(2a^2b^2+ab+a)_3$

TABLE A (*Continued*)

No.	Character Ia	Ib	IIa	IIb	IIIa	IIIb	IVa	IVb	V	VI	$a+b$	ab
100	−	+	+	+	−	−	+	−	+	+	ab of system 11	$(a)_3$
101	−	+	+	+	−	−	+	+	−	+	“ “ “ 12	ab of system 72
102	−	+	+	+	−	−	+	+	+	+	“ “ “ 13	“ “ “ 65
103	−	+	+	+	−	+	−	−	−	+	$a+b$ “ “ 33	$(ab)_3$
104	−	+	+	+	−	+	−	−	+	+	“ “ “ 34	$(a+b)_3$
105	−	+	+	+	−	+	−	+	−	+	ab “ “ 13	ab of system 68
106	−	+	+	+	−	+	−	+	+	+	$a+b$ “ “ 30	“ “ “ 104
107	−	+	+	+	−	+	+	−	−	+	“ “ “ 37	$(ab+2a+2b+2)_3$
108	−	+	+	+	−	+	+	−	+	+	“ “ “ 38	$(2ab^2+2a^2b+a+b)_3$
109	−	+	+	+	−	+	+	+	−	+	ab “ “ 13	ab of system 69
110	−	+	+	+	−	+	+	+	+	+	$a+b$ “ “ 30	$(2a^2b^2+2a^2b+2ab^2+ab)_3$
111	−	+	+	+	+	−	−	−	−	+	ab “ “ 16	ab of system 63
112	−	+	+	+	+	−	−	−	+	+	“ “ “ “	“ “ “ 64
113	−	+	+	+	+	−	−	+	−	+	“ “ “ 39	“ “ “ “
114	−	+	+	+	+	−	−	+	+	+	“ “ “ 38	$(ab^2)_3$
115	−	+	+	+	+	−	+	−	−	+	“ “ “ 18	ab of system 63
116	−	+	+	+	+	−	+	−	+	+	“ “ “ 16	“ “ “ 65
117	−	+	+	+	+	−	+	+	−	+	“ “ “ 39	“ “ “ 63
118	−	+	+	+	+	−	+	+	+	+	“ “ “ “	“ “ “ 65
119	−	+	+	+	+	+	−	−	−	+	“ “ “ 16	$(a+b+1)_2$
120	−	+	+	+	+	+	−	−	+	+	“ “ “ “	ab of system 68
121	−	+	+	+	+	+	−	+	−	+	“ “ “ 39	“ “ “ “
122	−	+	+	+	+	+	−	+	+	+	$a+b$ “ “ 44	$(a^2b+ab^2+a+b)_3$
123	−	+	+	+	+	+	+	−	−	+	$(ab)_3+0/[a, b; 2, 2]_3$	ab of system 104
124	−	+	+	+	+	+	+	−	+	+	ab of system 18	“ “ “ 119
125	−	+	+	+	+	+	+	+	−	+	“ “ “ 39	“ “ “ 69
126	−	+	+	+	+	+	+	+	+	+	“ “ “ 18	“ “ “ 78
127	+	+	−	−	−	−	−	−	+	+	“ “ “ 49	“ “ “ 49
128	+	+	−	−	−	−	−	+	+	+	“ “ “ “	“ “ “ 48
129	+	+	−	−	−	−	+	+	+	+	“ “ “ 50	“ “ “ 50
130	+	+	−	−	−	+	−	−	+	+	“ “ “ 48	“ “ “ 51
131	+	+	−	−	−	+	−	+	+	+	$(2a^2)_3$	$(0)_3$
132	+	+	−	−	−	+	+	−	+	+	$a+b$ of system 131	ab of system 53
133	+	+	−	−	−	+	+	+	+	+	ab “ “ 50	“ “ “ 51
134	+	+	−	−	+	+	−	−	+	+	“ “ “ 51	“ “ “ 59
135	+	+	−	−	+	+	−	+	+	+	“ “ “ 53	“ “ “ 131
136	+	+	−	−	+	+	+	+	+	+	“ “ “ 51	“ “ “ 51
137	+	+	−	+	−	−	−	−	+	+	“ “ “ 48	“ “ “ 64
138	+	+	−	+	−	−	−	+	+	+	“ “ “ 81	“ “ “ 72
139	+	+	−	+	−	−	+	−	+	+	“ “ “ 49	“ “ “ 65
140	+	+	−	+	−	−	+	+	+	+	“ “ “ 50	“ “ “ “
141	+	+	−	+	−	+	−	−	+	+	“ “ “ 48	“ “ “ 69
142	+	+	−	+	−	+	−	+	+	+	“ “ “ “	“ “ “ 68
143	+	+	−	+	−	+	+	−	+	+	“ “ “ 49	“ “ “ 69
144	+	+	−	+	−	+	+	+	+	+	“ “ “ 47	“ “ “ “
145	+	+	−	+	+	−	−	−	+	+	“ “ “ 59	“ “ “ 64

TABLE B (*Continued*)

No.	Character Ja	Ib	IIa	IIb	IIIa	IIIb	IVa	IVb	V	VI	$a+b$	ab
146	+	+	−	+	+	−	−	+	+	+	ab of system 85	ab of system 72
147	+	+	−	+	+	−	+	−	+	+	" " " 51	" " " 65
148	+	+	−	+	+	−	+	+	+	+	" " " 85	" " " 100
149	+	+	−	+	+	+	−	−	+	+	" " " 52	" " " 69
150	+	+	−	+	+	+	−	+	+	+	" " " 85	$(2ab+a+b)_3$
151	+	+	−	+	+	+	+	−	+	+	" " " 59	ab of system 69
152	+	+	−	+	+	+	+	+	+	+	" " " 51	" " " "
153	+	+	+	+	−	−	−	−	−	+	" " " 64	" " " 63
154	+	+	+	+	−	−	−	−	+	+	" " " "	" " " 64
155	+	+	+	+	−	−	−	+	−	+	" " " 99	" " " 72
156	+	+	+	+	−	−	−	+	+	+	" " " 65	" " " 64
157	+	+	+	+	−	−	+	+	−	+	" " " 72	" " " 72
158	+	+	+	+	−	−	+	+	+	+	" " " 65	" " " 65
159	+	+	+	+	−	+	−	−	−	+	" " " 63	" " " 69
160	+	+	+	+	−	+	−	−	+	+	" " " 64	" " " 68
161	+	+	+	+	−	+	−	+	−	+	" " " "	" " " 69
162	+	+	+	+	−	+	−	+	+	+	" " " 65	" " " 68
163	+	+	+	+	−	+	+	−	−	+	$(2a^2b+ab+2a^2+2a)_3$	" " " 103
164	+	+	+	+	−	+	+	−	+	+	ab of system 114	" " " 108
165	+	+	+	+	−	+	+	+	−	+	" " " 72	$(a^2b^2+2a^2b+2ab^2+2ab)_3$
166	+	+	+	+	−	+	+	+	+	+	" " " 65	ab of system 69
167	+	+	+	+	+	+	−	−	−	+	" " " 119	" " " "
168	+	+	+	+	+	+	−	−	+	+	" " " 68	" " " 68
169	+	+	+	+	+	+	−	+	−	+	" " " 104	" " " 103
170	+	+	+	+	+	+	−	+	+	+	" " " 68	" " " 69
171	+	+	+	+	+	+	+	+	−	+	" " " 69	" " " "
172	+	+	+	+	+	+	+	+	+	+	" " " 67	" " " "

5. **Remarks on the systems.** In the proof-systems denying postulate Ia, the elements a and b for which $a+b$ does not exist are respectively the values of m and n which occur in the symbol $[a, b; m, n]_p$. Similarly, for postulate Ib.

The verifications will often be more easily effected if tables are constructed corresponding to the modular expressions in question.

UNIVERSITY OF CALIFORNIA,
BERKELEY, CALIF.

CYCLIC FIELDS OF DEGREE EIGHT*

BY
A. ADRIAN ALBERT

1. **Introduction.** Let F be any non-modular field, C be an algebraic extension of degree n of F. Then $C=F(x)$ is the field of all rational functions with coefficients in F of a root x of an equation $\phi(\omega)=0$ which has coefficients in F, degree n, and transitive group G for F.

The problem of the construction of all equations of degree n and group G is evidently equivalent to the problem of the construction of all corresponding fields C. Moreover the construction of a set of canonical equations $\psi(\omega)=0$ with the property that every $C=F(x)$ of degree n and group G is equal to an $F(y)$ defined by a $\psi(\omega)=0$ provides a solution of both problems.

One of the most important problems in the algebraic theory of fields is the construction of all cyclic fields of degree n over F. This is the case where G consists of the n distinct powers S^i $(i=0, 1, \cdots, n-1)$ of a single substitution S. In this case G is also the group of all automorphisms of C. Moreover this problem has been reduced to the case $n=p^e$, p a prime.

Cyclic fields of degree 2, 2^2 have been constructed.† In the present paper we shall use purely algebraic methods to construct all cyclic fields of degree $2^3=8$‡ over any non-modular field F.

2. **General theory of cyclic fields.** Let F be any non-modular field and let C be a cyclic field of degree n over F. Then if

$$n = p_1^{e_1}\cdot p_2^{e_2}\cdots p_t^{e_t}, \tag{1}$$

where the p_i are distinct primes, it is well known that C is the direct product

$$C = C^{(1)} \times C^{(2)} \times \cdots \times C^{(t)} \tag{2}$$

of cyclic fields $C^{(i)}$ of degree $p_i^{e_i}$ over F. Conversely every direct product (2) is a cyclic field of degree n over F. It is thus certain that the problem of constructing all cyclic fields of degree n over F is equivalent to the corresponding problem for the case $n=p^e$.

* Presented to the Society, February 25, 1933; received by the editors December 19, 1932.

† Cf. §2.

‡ Cyclic fields of degree eight have been considered by F. Mertens in the Wiener Sitzungsberichte, vol. 125 (1916), pp. 741–831. But he considered algebraic number fields, used the arithmetic theory of ideals, and did not give very explicit results. His method is not at all applicable to the case we are considering (where F is a general field). Moreover I believe the results obtained here are more explicit and give a more definite construction for C even for the cases considered by Mertens.

Let then $C=C_e$ have degree $n=p^e$, p a prime. It is well known that we may define a chain of fields

$$C_e > C_{e-1} > \cdots > C_1 > C_0 = F, \tag{3}$$

where C_i is cyclic of degree p^i over F, cyclic of degree p over C_{i-1}. In fact let S be the automorphism of C generating its group G of automorphisms. Then this group of order n is given by

$$G_e = (I, S, S^2, \cdots, S^{n-1}), S^n = I. \tag{4}$$

But if $T=S^{p^{e-1}}$ then

$$H = (I, T, T^2, \cdots, T^{p-1}) \tag{5}$$

is an invariant sub-group of G of index p^{e-1} defining a sub-field C_{e-1} of degree p^{e-1} and with Galois group

$$G_{e-1} = (I, \sigma, \sigma^2, \cdots, \sigma^{m-1}), m = p^{e-1}, \tag{6}$$

isomorphic with G_e but with $T=S^m$ in G corresponding to the identity of G_{e-1}.

We may now consider every C_e as a cyclic field of degree p over a cyclic field C_{e-1} of degree p^{e-1} to obtain some of the properties of C_e. But if C_e is cyclic of degree p over C_{e-1} which is cyclic of degree p^{e-1} over F, then it is not necessarily true that C_e is cyclic over F. Thus we shall also require a consideration of further properties.

We are interested here only in the case $p=2$. Let C be a cyclic field of degree $n=2^e$, $e>1$ over F, and let D be its uniquely defined sub-field of degree $m=2^{e-1}$. Then C is a quadratic field over D,

$$C = D(x), \ x^2 = a \text{ in } D, \tag{7}$$

where 1, x are linearly independent with respect to D. The substitution S generating the cyclic group of C has order n and D consists of all quantities d of C such that

$$d^{S^m} = d \qquad (m = 2^{e-1}). \tag{8}$$

For convenience of notation we shall write

$$c^{S^k} = c^{(k)}, \tag{9}$$

so that $c^{(m)}=c$ whenever c is in D but not otherwise. Then

$$[x^{(m)}]^2 = a^{(m)} = a = x^2,$$

and $x^{(m)} = \pm x$. But x is not in D. Hence

$$x^{(m)} = -x. \tag{10}$$

In particular let $x' = \alpha + \beta x$ where α and β are in D. Then $(x')^2 = a' = \alpha^2 + \beta^2 a + 2\alpha\beta x$. But a' is in D so that $2\alpha\beta = 0$. If $\beta = 0$ then $x' = \alpha$ is in D and $(x')^{(n-1)} = x = \alpha^{(n-1)}$ is in D, a contradiction. Hence $\beta \neq 0$, $\alpha = 0$ and

$$x' = \beta x, \qquad \beta \text{ in } D. \tag{11}$$

It is obvious that $D(x) = D(bx)$ for every non-zero b of D. Hence all of the above properties as well as those we may derive later will hold for any bx taken as the quantity generating C, a quadratic field over D.

We shall assume first that $n = 2$. Then $D = F$ and, since x is not in D, the field $F(x) = D(x) = C$ is a quadratic field over F generated by x. Let next

$$m = 2^{e-1} = 2g, \; g \geqq 1,$$

so that $n \geqq 4$. Then $D = K(y)$, $y^2 = \alpha$ in K, is a quadratic field over the field K of all quantities k of C such that

$$k^{(g)} = k.$$

The field $F(x)$ is a sub-field of $C = D(x)$. But x is not in $D \geqq F(x^2) = F(a)$ so that the degree of $F(x)$ is $2h$ where h is the degree of $F(x^2)$. Hence $F(x) = C$ if and only if $F(a) = D$.

Suppose that $F(a) < D$. Then a is in a proper sub-field of D. But D is cyclic and its maximal proper sub-field K contains every proper sub-field of D. Hence a is in K, $a^{(g)} = a$, $[x^{(g)}]^2 = a^{(g)} = x^2$. Then $x^g = \pm x$, $x^{(m)} = [x^{(g)}]^{(g)} = x$, a contradiction. Hence $F(a) = D$ and we have

Theorem 1. *Let C be a cyclic field of degree $n = 2m$ over F, $C = D(x)$ where D is a cyclic sub-field of C of degree $m = 2^{e-1}$ over F so that x may be chosen so that*

$$x^2 = a \text{ in } D.$$

Then $x' = \beta \cdot x$ where β is in D and has the property that $x^{(m)} = -x$. Moreover this latter property implies that $F(x) = C$, $F(a) = D$.

Suppose that $x_0 = bx$, $b \neq 0$ in D. Then $x_0^2 = b^2x^2 = b^2a$ is in D, $x_0^{(m)} = b^{(m)}x^m = -bx = -x_0$. By Theorem 1, $C = F(x) = D(x) = D(x_0) = F(bx)$.

Theorem 2. *Let $x_0 = bx$ where $b \neq 0$ is in D. Then $F(x) = F(x_0) = C$.*

The condition $x' = \beta \cdot x$ imposes two restrictions on β. The first is obviously $x'^2 = \beta^2 \cdot x^2 = (x^2)' = a' = \beta^2 \cdot a$, a necessary and sufficient condition that x' shall actually equal $\beta \cdot x$. Next we must have $x^{(m)} = -x$. But

$$x'' = (\beta x)' = \beta'\beta x, \cdots, x^{(k)} = [x^{(k-1)}]' = [\beta^{(k-1)}\beta^{(k-2)} \cdots \beta'\beta]x,$$

and

$$x^{(m)} = [\beta^{(m-1)}\beta^{(m-2)} \cdots \beta'\beta]x = -x,$$

so if we write

$$N_D(\beta) = \beta\beta'\beta'' \cdots \beta^{(m-1)},$$

then it follows from $x^{(m)} = -x$ that

$$N_D(\beta) = -1. \tag{12}$$

Conversely let D be cyclic of degree $m = 2^{e-1}$ over F and let a, β satisfy (12). Then the field $D(x)$ defined by a root x of $x^2 = a$ is a quadratic field over D if and only if a is not the square of any quantity of D. But if $a = c^2$, c in D, then $\beta^2 = (a')(a)^{-1} = (c'c^{-1})^2$ so that

$$\beta = \pm (c')(c)^{-1}. \tag{13}$$

But m is even and

$$N_D(\beta) = (\pm 1)^m N_D\left(\frac{c'}{c}\right) = (\pm 1)^m = 1, \tag{14}$$

a contradiction of the first equation of (12). Hence $D(x)$ has degree $n = 2m$. Also if we define $x' = \beta \cdot x$ then (12) implies that $x^{(m)} = -x$ so that we have defined a self correspondence of $C = D(x)$

$$c + dx \longleftrightarrow c' + d'x' = c' + d'\beta x, \tag{15}$$

for every c and d of D, $c + dx$ of C. This correspondence is evidently preserved under addition, subtraction, multiplication and division and is an automorphism of C if and only if $x'^2 = a'$ which is satisfied by (12). Hence (15) is an automorphism S of C and, since S^m is an automorphism of C in which x corresponds to $-x$ the order of S is $n = 2m$ and C is a cyclic field. Obviously D is the set of all quantities of C unaltered by S^m. By Theorem 1, $C = F(x)$ and we have proved

THEOREM 3. *Let D be cyclic of degree 2^{e-1} over F with generating automorphism*

$$d \longleftrightarrow d',$$

for every d of D. Then D is the unique sub-field of degree m of a cyclic field C of degree $n = 2m$ if and only if there exist quantities $\beta \neq 0$, $a \neq 0$ in D, such that

$$\beta^2 = \frac{a'}{a}, \quad N_D(\beta) = -1. \tag{16}$$

Moreover every solution of (16) *defines a cyclic field of degree n over F*

$$C = F(x),\ x^2 = a \text{ in } D,\ x \longleftrightarrow x' = \beta \cdot x, \tag{17}$$

as generating automorphism, so that D is the set of all quantities d of C such that $d^{(m)} = d$.

The case $m=1$, $n=2$ is trivial so that we shall assume henceforth that $n>2$, $n=4g=2m$. Then (16) implies that

$$(18)\ \beta^2 = \frac{a'}{a},\quad \beta'^2 = \frac{a''}{a'},\ \cdots,\ (\beta^{(i-1)})^2 = \frac{a^{(i)}}{a^{(i-1)}},\ \cdots,\ (\beta^{(g-1)})^2 = \frac{a^{(g)}}{a^{(g-1)}},$$

and hence that

$$(19)\qquad [\beta\beta' \cdots \beta^{(g-1)}]^2 = \frac{a^{(g)}}{a}\,.$$

Then if

$$(20)\qquad y = a\beta\beta' \cdots \beta^{(g-1)},$$

equation (19) implies

$$(21)\qquad y^2 = aa^{(g)}.$$

But D is a quadratic field $K(d)$, d^2 in K, over a cyclic field K of degree g over F. Moreover

$$(22)\qquad k^{(g)} = k$$

for every k of K. Since $m=2g$, $a^{(m)}=a$, we have $[aa^{(g)}]^{(g)}=a^{(g)}a$ so that y^2 is in K. Also $yy^{(g)}=aa^{(g)}[\beta\beta'\cdots\beta^{(g-1)}][\beta^{(g)}\cdots\beta^{(m-1)}]=aa^{(g)}N_D(\beta)=-aa^{(g)}$ $=-y^2$, $y^{(g)}=-y$, and y is not in K. But then y generates D, a quadratic field over K, and

$$(23)\qquad D = K(y),\ y^2 = \alpha \text{ in } K.$$

The field $D=K(y)$ is a quadratic field over K which is cyclic of degree g over F. By Theorem 3 there exist quantities $\alpha=y^2$ in K, $\gamma=y'y^{-1}$ in K, such that

$$(24)\qquad \gamma^2 = \frac{\alpha'}{\alpha},\quad N_K(\gamma) \equiv \gamma\gamma' \cdots \gamma^{(g-1)} = -1,$$

and, by this same theorem, $D=F(y)$. Hence

$$(25)\qquad C = F(x),\ x^2 = a \text{ in } D,\ D = F(y),\ y^2 = \alpha \text{ in } K,$$

$$(26)\qquad x' = \beta x,\ y' = \gamma y,\ a = \frac{y}{\beta\beta' \cdots \beta^{(g-1)}}\,.$$

We now wish

$$(27)\qquad \beta^2 = \frac{a'}{a} = \frac{y'}{\beta'\beta'' \cdots \beta^{(g)}}\ \frac{\beta\beta' \cdots \beta^{(g-1)}}{y} = \frac{y'}{y}\ \frac{\beta}{\beta^{(g)}}\,.$$

But (27) is equivalent to

$$\frac{y'}{y} = \beta\beta^{(g)}, \tag{28}$$

that is, $\gamma = \beta\beta^{(g)}$.

Conversely, let $y^2 = \alpha$ in K, $\gamma^2 = \alpha'\alpha^{-1}$, $N_K(\gamma) = -1$, so that, by Theorem 3, $K(y) = F(y)$ is a cyclic field of degree $2g$ over F. Let also a be defined by the third equation of (26), and β be in D and satisfy

$$\gamma = \beta\beta^{(g)}. \tag{29}$$

Then

$$N_D(\beta) = [\beta\beta' \cdots \beta^{(g-1)}][\beta\beta' \cdots \beta^{(g-1)}]^{(g)} = [\beta\beta^{(g)}][\beta\beta^{(g)}]' \cdots [\beta\beta^{(g)}]^{(g-1)}$$
$$= \gamma\gamma' \cdots \gamma^{(g-1)} = N_K(\gamma) = -1.$$

Also

$$\frac{a'}{a} = \frac{y'}{\beta'\beta'' \cdots \beta^{(g)}} \frac{\beta\beta' \cdots \beta^{(g-1)}}{y} = \frac{\gamma\beta}{\beta^{(g)}} = \frac{\gamma}{\beta\beta^{(g)}} \cdot \beta^2 = \beta^2,$$

as desired. We are now in a position to prove

THEOREM 4. *Let $n = 4g = 2^e$ and let K be a cyclic field of degree g over F with automorphism $k \longleftrightarrow k'$ for every k of K. Then K is the unique sub-field of degree g of a cyclic field C of degree n over F if and only if there exist quantities $\alpha \neq 0$, $\gamma \neq 0$, β_1, β_2 in K satisfying*

$$\gamma^2 = \frac{\alpha'}{\alpha}, \quad N_K(\gamma) = -1, \gamma = \beta_1^2 - \beta_2^2\alpha. \tag{30}$$

Every solution of (30) *defines a cyclic field $F(x) = C$ with*

$$y^2 = \alpha, \beta = \beta_1 + \beta_2 y, x^2 = a = \frac{y}{\beta\beta' \cdots \beta^{(g-1)}}, \tag{31}$$

and with generating automorphism S given by

$$c = c_1 + c_2 y + (c_3 + c_4 y)x \longleftrightarrow c' = c_1' + c_2'\gamma y + (c_3' + c_4'\gamma y)\beta x, \tag{32}$$

for every c_1, c_2, c_3, c_4 of K, c of C, so that

$$x' = \beta x, \; y' = \gamma y. \tag{33}$$

We obviously also will have

COROLLARY 1. *In Theorem 4 the field K is the field of all quantities of C unaltered by S^{σ}, the field $D=F(y)=K(y)$ is the field of all quantities of C unaltered by S^{m}.*

For we need only notice in the above that, since β is in D, $\beta=\beta_1+\beta_2 y$ where β_1 and β_2 are in K. We have then merely replaced the condition $\gamma=\beta\beta^{(\sigma)}$ by the equivalent condition $\gamma=\beta_1^2-\beta_2^2\alpha$ of (30).

We shall now obtain some important restrictions which it is possible to impose on β. Suppose first that $n=4$, $m=2$. Then $K=F$, $N_K(\gamma)=\gamma=-1$ is in F, and (30) becomes merely $\beta_1^2-\beta_2^2\alpha=-1$. If $\beta_1=0$ then $\beta_2\neq 0$, $\alpha=(\beta_2^{-1})^2$ which is impossible if D is a quadratic field over F. Hence for this case $\beta_1\neq 0$.

There exists the possibility in the above theorem that $\beta_1\beta_2=0$. We shall be able to restrict β so that all fields C are obtained yet $\beta_1\beta_2\neq 0$.

By Theorem 2 if $b=b_1+b_2 y$, $b_1 b_2\neq 0$, b_1 and b_2 in K, then $x_0=bx$ also generates $F(x)$ and satisfies

$$(34)\qquad x_0^2=a_0=\frac{y_0}{\beta_0\beta_0'\cdots\beta_0^{(\sigma-1)}},\quad x_0'=\beta_0 x_0,\quad y_0^2=\alpha_0,\quad y_0'=\gamma_0 y_0,$$

with

$$(35)\qquad \gamma_0^2=\frac{\alpha_0'}{\alpha_0},\quad N_K(\gamma_0)=-1,\quad \beta_0=\beta_{10}+\beta_{20}y_0,\quad \beta_{10}^2-\beta_{20}^2\alpha_0=\gamma_0.$$

But

$$(36)\qquad x_0'=(bx)'=b'\beta x=\beta_0 x_0=\beta_0 bx,$$

$$(37)\qquad \begin{aligned}\beta_0=\frac{b'}{b}\beta&=\frac{b_1'+b_2'\gamma y}{b_1+b_2 y}\beta=(b_1'+b_2'\gamma y)(b_1-b_2 y)e\\ &=[(b_1'b_1-b_2'b_2\alpha\gamma)+(b_2'\gamma b_1-b_1'b_2)y]e,\end{aligned}$$

where

$$(38)\qquad e=(b_1^2-b_2^2\alpha)^{-1}\beta$$

is either in K or a multiple of y by a quantity of K according as $\beta_2=0$ or $\beta_1=0$. But then $\beta_{10}\beta_{20}=0$ if and only if

$$(39)\qquad b_2'\gamma b_1-b_1'b_2=0,\ \text{or}\ b_1'b_1-b_2'b_2\alpha\gamma=0.$$

Suppose first that $b_2'\gamma b_1-b_1'b_2=0$. Then since $b_1b_2\neq 0$,

$$(40)\qquad \gamma=\frac{b_1'b_2}{b_1b_2'},\quad \gamma'=\frac{b_1''b_2'}{b_1'b_2''},\ \cdots,\ \gamma^{(\sigma-1)}=\frac{b_1^{(\sigma)}b_2^{(\sigma-1)}}{b_1^{(\sigma-1)}b_2^{(\sigma)}}$$

and

$$(41)\qquad -1 = N_K(\gamma) = \frac{b_1' b_2}{b_1 b_2'}\cdot\frac{b_1'' b_2'}{b_1' b_2''}\cdots\frac{b_1^{(\sigma)} b_2^{(\sigma-1)}}{b_1^{(\sigma-1)} b_2^{(\sigma)}} = \frac{b_1^{(\sigma)}}{b_1}\cdot\frac{b_2}{b_2^{(\sigma)}} = 1,$$

since $b_1 = b_1^{(\sigma)}$ and $b_2 = b_2^{(\sigma)}$ are in K, a contradiction. Hence $b_2'\gamma b_1 - b_1' b_2 \neq 0$.

We have then proved that if $\beta_{20}\beta_{10} = 0$ then $b_1 b_1' = b_2 b_2' \alpha\gamma$. If $n = 4$ then $m = 2$ and we have already shown that $\beta_1 \neq 0$. Hence $\beta_{10} \neq 0$ and hence $\beta_2 = \beta_{20} = 0$. But the coefficient of y in β_0 when $e \neq 0$ is in $F = K$, that is, $\beta_2 = 0$, is $\beta_{20} = d(b_2'\gamma b_1 - b_1' b_2) \neq 0$ as we have shown. It remains only to consider the case $n > 4$, $m > 2$, $b_1 b_1' = \alpha\gamma b_2 b_2'$.

Let $y_0 = b_2 b_1^{-1} y$. Then $F(y_0) = F(y)$,

$$(42)\qquad y_0 y_0' = b_2 b_2' (b_1 b_1')^{-1}\alpha\gamma = 1,\ y_0' = (y_0)^{-1},\ y_0'' = y_0.$$

But the automorphism S of $D = F(y_0)$ replacing y_0 by y_0' has order m. Hence $m = 2$, a contradiction. We have proved

THEOREM 5. *Every cyclic field $F(x)$ of degree $n = 2^e$ over F with K as cyclic sub-field is generated by an x of Theorem 4 with $\beta_1\beta_2 \neq 0$ in* (30).

3. **Cyclic quartic fields.** Let $n = 4$ so that $K = F$, $g = 1$, γ and α are in F. Then $N_K(\gamma) = \gamma = -1$, $\beta_1^2 - \beta_2^2\alpha = -1$ for $\beta_1 \neq 0$, $\beta_2 \neq 0$ in F. Put $\epsilon = \beta_1^{-1}$ and obtain $-\epsilon^2 = 1 - (\beta_2\beta_1^{-1})^2\alpha$, whence if $u = \beta_2\epsilon y$ then $F(u) = F(y)$ and $u^2 = \beta_2^2\epsilon^2 y^2 = (\beta_2\beta_1^{-1})^2\alpha$,

$$(43)\qquad u^2 = 1 + \epsilon^2 = \tau \text{ in } F,\ \beta = \frac{1+u}{\epsilon},$$

since $\epsilon\beta = \epsilon(\beta_1 + \beta_2 y) = 1 + u$. Also

$$(44)\qquad x^2 = a = \frac{y}{\beta} = \frac{\beta_2\epsilon y}{\beta_2(1+u)} = \frac{u}{\beta_2(1+u)}\cdot\frac{1-u}{1-u} = \nu(u - \tau),$$

where $\nu = (-\beta_2\epsilon^2)^{-1} \neq 0$ is in F. We have therefore proved the well known result

THEOREM 6. *Every cyclic field $F(x)$ of degree four over F is generated by a quantity x satisfying*

$$(45)\qquad x^2 = \nu(u - \tau),\ x' = \frac{1+u}{\epsilon}x,\ u^2 = \tau = 1 + \epsilon^2,$$

where ϵ and $\nu \neq 0$ are in F and τ is not the square of any quantity of F.

4. **Cyclic fields of degree eight.** Let now $n = 8$, $m = 4$, $g = 2$. Then $F(y)$ is a cyclic quartic field. We wish $\beta = \beta_1 + \beta_2 y$ with $y^2 = \alpha$ in K, γ, β_1, β_2 in K and $\gamma\gamma' = -1$,

$$(46)\qquad \beta_1^2 - \beta_2^2\alpha = \gamma,\ \beta_1\beta_2 \neq 0.$$

Let

$$\delta = \beta y = \beta_2\alpha + \beta_1 y = \delta_1 + y_0, \tag{47}$$

where

$$F(y_0) = F(y),\ y_0 = \beta_1 y,\ \delta_1 = \beta_2\alpha \text{ in } K. \tag{48}$$

Then

$$\beta = y^{-1}\delta = (\beta_1 y_0^{-1})\delta,\ (\beta_2\alpha)^2 - \beta_1^2\alpha = -\alpha\gamma, \tag{49}$$

so that, since

$$y_0^2 = \alpha_0 = \beta_1^2\alpha,\ y_0' = \gamma_0 y_0, \tag{50}$$

we have

$$\beta = (\beta_1\alpha_0^{-1})\delta y_0 = (\beta_1\alpha_0^{-1})(\alpha_0 + \delta_1 y_0). \tag{51}$$

Also $\beta\beta'\beta''\beta''' = -1$ and hence

$$a = \frac{y}{\beta\beta'} = -y(\beta\beta')'' = -y\left(\frac{\delta\delta'}{yy'}\right)'' = -\frac{y(\delta\delta')''}{\alpha\gamma} \tag{52}$$

$$= -\frac{y_0}{\beta_1}(\beta_1\beta_1')\frac{(\delta\delta')''}{\alpha_0\gamma_0} = \frac{\beta_1'(\delta\delta')''}{-\alpha_0\gamma_0}y_0,$$

where

$$\gamma_0\gamma_0' = -1,\ \gamma_0^2 = \frac{\alpha_0'}{\alpha_0},\ \delta_1^2 - \alpha_0 = -\lambda^{-1}\alpha_0\gamma_0,\ \lambda = \beta_1\beta_1'. \tag{53}$$

Suppose that γ_0 is in F. Then $\gamma_0\gamma_0' = -1$ gives $\gamma_0^2 = -1$, $\gamma_0 = i = (-1)^{1/2}$. Also $\alpha_0' = -\alpha_0$ and if $K = F(u)$ we may take $\alpha_0 = u$. Then the solution of (46) is equivalent to $\delta_1^2 - \alpha_0 = -\lambda^{-1}\alpha_0\gamma_0$ where λ is in F and hence to the solution of $\delta_1^2 = u(1-\lambda^{-1}i)$. But if $\delta_1 = \xi_1 + \xi_2 u$ this implies that $u^2 = \tau$ in F, $\xi_1^2 + \xi_2^2\tau + 2\xi_1\xi_2 u = u(1-\lambda^{-1}i)$, $\xi_1^2 + \xi_2^2\tau = 0$ and $\tau = -(\xi_1\xi_2^{-1})^2 = (i\xi_1\xi_2^{-1})^2$, a contradiction of our hypothesis that $F(u)$ is a quadratic field.

Hence γ_0 is not in F and the hypothesis $\beta_2 \neq 0$ of §3 is satisfied for $F(y_0)$. But then

$$y_0^2 = \alpha_0 = \nu(u - \tau),\ \frac{y_0'}{y_0} = \gamma_0 = \frac{1+u}{\epsilon},\ u^2 = \tau = 1 + \epsilon^2. \tag{54}$$

Also $-\alpha_0\gamma_0 = -\nu\epsilon^{-1}(u-\tau)(u+1) = \nu\epsilon^{-1}u(\tau - 1)$; that is, since $\tau - 1 = \epsilon^2$,

$$-\alpha_0\gamma_0 = \nu\epsilon u. \tag{55}$$

We may now complete our computation (52) of a. We use

$$(\delta\delta')''y_0 = [(\delta_1 + y_0)(\delta_1' + y_0')]''y_0 = (\delta_1 - y_0)(\delta_1' - \gamma_0 y_0)y_0$$
$$= (-\alpha_0 + \delta_1 y_0)(\delta_1' - \gamma_0 y_0) = -(\alpha_0\delta_1' + \delta_1\alpha_0\gamma_0) + (\delta_1\delta_1' + \alpha_0\gamma_0)y_0.$$

Hence

$$(56) \qquad a = \frac{\beta_1' u}{\nu\epsilon\tau}[\nu\epsilon u\delta_1' - \nu(u - \tau)\delta_1 + (\delta_1\delta_1' - \nu\epsilon u)y_0],$$

where

$$(57) \qquad \delta_1 = \xi_1 + \xi_2 u, \delta_1' = \xi_1 - \xi_2 u, \beta_1 = \xi_3 + \xi_4 u.$$

Also (51) gives $\beta = \beta_1(-\alpha_0\gamma_0)^{-1}(-\alpha_0\gamma_0 - \delta_1\gamma_0 y_0) = \beta_1(\nu\epsilon u)^{-1}(\nu\epsilon u - \delta_1\gamma_0 y_0)$ and hence

$$58) \qquad \beta = \frac{\beta_1}{\nu\epsilon\tau}\left[\nu\epsilon\tau - \frac{\delta_1}{\epsilon}(u + \tau)y_0\right].$$

We have proved

THEOREM 7. *Every cyclic field $F(x)$ of degree eight over F is generated by a quantity x satisfying*

$$(59) \qquad x^2 = a, \; x' = \beta x,$$

with a and β given by (54), (56), (57), (58) *such that $\nu \neq 0$ in F, $\delta_1 \neq 0$, $\beta_1 \neq 0$, and if*

$$(60) \qquad \lambda = \xi_3^2 - \xi_4^2\tau,$$

then

$$(61) \qquad \delta_1^2 = \alpha_0 - \lambda^{-1}\alpha_0\gamma_0 = \nu(u - \tau + \lambda^{-1}\epsilon u).$$

The quantity $\delta_1^2 = \xi_1^2 + \xi_2^2\tau + 2\xi_1\xi_2 u$, so that (61) is equivalent to

$$(62) \qquad -\nu\tau = \xi_1^2 + \xi_2^2\tau, \; 2\xi_1\xi_2 = \nu(1 + \lambda^{-1}\epsilon).$$

But then $-2\xi_1\xi_2\tau = (-\nu\tau)(1 + \lambda^{-1}\epsilon) = (1 + \lambda^{-1}\epsilon)(\xi_1^2 + \xi_2^2\tau)$, so that, since $\nu \neq 0$, equation (61) is equivalent to

$$(63) \qquad \nu = (-\tau)^{-1}(\xi_1^2 + \xi_2^2\tau) \neq 0, \; 1 + \lambda^{-1}\epsilon = \frac{-2\xi_1\xi_2\tau}{\xi_1^2 + \xi_2^2\tau}.$$

The first equation of (63) will be taken to determine ν. The second equation becomes

$$(64) \qquad -\epsilon = \lambda\left[1 + \frac{2\xi_1\xi_2\tau}{\xi_1^2 + \xi_2^2\tau}\right] = (\xi_3^2 - \xi_4^2\tau)\frac{(\xi_1^2 + \xi_2^2\tau + 2\xi_1\xi_2\tau)}{\xi_1^2 + \xi_2^2\tau},$$

to be solved for $\xi_1^2+\xi_2^2\tau\neq 0$. But $\xi_1^2+\xi_2^2\tau+2\xi_1\xi_2\tau=(\xi_1+\xi_2\tau)^2+\xi_2^2\tau(1-\tau)$ $=(\xi_1+\xi_2\tau)^2-(\xi_2\epsilon)^2\tau$. Hence if

$$k = \eta_1 + \eta_2 u = \frac{(\xi_3 + \xi_4 u)(\xi_1 + \xi_2\tau + \xi_2\epsilon u)}{\xi_1^2 + \xi_2^2\tau}, \tag{65}$$

where η_1 and η_2 are then explicitly determined in terms of ξ_1, ξ_2, ξ_3, ξ_4, then

$$kk' = \frac{(\xi_3^2 - \xi_4^2\tau)(\xi_1^2 + \xi_2^2\tau + 2\xi_1\xi_2\tau)}{(\xi_1^2 + \xi_2^2\tau)^2} = -\frac{\epsilon}{\xi_1^2 + \xi_2^2\tau}, \tag{66}$$

so that, since $kk'=\eta_1^2-\eta_2^2\tau$,

$$-\epsilon = (\xi_1^2 + \xi_2^2\tau)(\eta_1^2 - \eta_2^2\tau) \neq 0, \tag{67}$$

where we use $\tau=1+\epsilon^2\neq 1$.

Conversely let $\epsilon\neq 0$ satisfy (67) and define k by $k=\eta_1+\eta_2 u$. Define

$$\beta_1 = \frac{(\xi_1^2 + \xi_2^2\tau)k}{\xi_1 + \xi_2\tau + \xi_2\epsilon u} = \xi_3 + \xi_4 u, \quad r = \xi_1 + \xi_2\tau + \xi_2\epsilon u, \tag{68}$$

where β_1 exists since $\xi_1^2+\xi_2^2\tau\neq 0$ and hence $r\neq 0$. Then we have

$$-\epsilon = kk'(\xi_1^2 + \xi_2^2\tau) = \frac{\beta_1\beta_1' rr'}{\xi_1^2+\xi_2^2\tau} = \frac{(\xi_3^2 - \xi_4^2\tau)(\xi_1^2 + \xi_2^2\tau + 2\xi_1\xi_2\tau)}{\xi_1^2 + \xi_2^2\tau}, \tag{69}$$

and (64) will be satisfied. Moreover if we define ν by $(63)_1$ then (61) will be satisfied. Also $\tau=1+\epsilon^2$ must not be the square of any quantity of F if $F(u)$, $u^2=\tau$, is a quadratic field over F as we are supposing. We have proved

THEOREM 8. *The solution of* (61) *is equivalent to the determination of* ν *by*

$$\nu = (-\tau)^{-1}(\xi_1^2 + \xi_2^2\tau), \tag{70}$$

and the solution of

$$-\epsilon = (\eta_1^2 - \eta_2^2\tau)(\xi_1^2 + \xi_2^2\tau) \tag{71}$$

for ϵ, η_1, η_2, ξ_1, ξ_2 *in* F *and such that* $\tau=1+\epsilon^2$ *is not the square of any quantity of* F.

5. **The formulas for α_0, γ_0, a, β.** We have seen how every cyclic field $F(x)$ of degree eight over F is generated by a quantity x such that $x^2=a$, $x'=\beta x$ where a and β are given by (56), (58), (54), (57) as soon as ν, ϵ, $\tau=1+\epsilon^2$, $\beta_1=\xi_3+\xi_4 u$, $\delta_1=\xi_1+\xi_2 u$ have been determined to satisfy (61). We have also shown that the solution of (61) is equivalent to (70) and the solution of the equation (71) with variables in F. Hence we have merely to solve (71), obtaining formulas with parameters for ϵ, η_1, η_2, ξ_1, ξ_2, obtain formulas for

ξ_3 and ξ_4 by the use of (68), and by the substitution of values so obtained in (54), (56), (57), (58) obtain explicit $a, \beta, \alpha_0, \gamma_0$. But the formulas so obtained would be undesirable because of complexity. Hence we shall confine our further work to a consideration of the only remaining non-trivial part of our problem, the solution of (71). Explicit fields of degree eight may then be obtained by carrying out the above work of substitution for every special case.

6. **The case i in F.** Suppose that F contains a quantity i such that $i^2=-1$. Then if $\tau=1+\epsilon^2$, ϵ in F, we wish to solve $-\epsilon=(\xi_1^2+\xi_2^2\tau)(\eta_1^2-\eta_2^2\tau)$ for $\xi_1, \xi_2, \eta_1, \eta_2$ in F and τ not the square of any quantity of F. Let

$$k_1 = \xi_1 + \xi_2 iu, \quad k_2 = \eta_1 + \eta_2 u, \tag{72}$$

so that k_1 is in $F(u)$, $u^2=1+\epsilon^2$, k_2 is in $F(u)$. Then if

$$k_3 = k_1 k_2 = \lambda + \mu u, \quad \lambda, \mu \text{ in } F, \tag{73}$$

we have

$$\lambda^2 - \mu^2\tau = -\epsilon, \tag{74}$$

since if $k_3'=\lambda-\mu u$ then $k_3k_3'=k_1k_1'\cdot k_2k_2'=[\xi_1^2-(\xi_2 i)^2\tau][\eta_1^2-\eta_2^2\tau]=(\xi_1^2+\xi_2^2\tau)(\eta_1^2-\eta_2^2\tau)=-\epsilon$.

Conversely let λ, μ be a solution of (74). Then if k_3 is defined by (73) we have

$$k_1 = (\xi_1 + \xi_2 iu) = \frac{k_3}{k_2} = \frac{\lambda + \mu u}{\eta_1 + \eta_2 u} = \frac{(\lambda\eta_1 - \eta_2\mu\tau)}{\eta_1^2 - \eta_2^2\tau} + \frac{\mu\eta_1 - \lambda\eta_2}{\eta_1^2 - \eta_2^2\tau}u,$$

so that

$$\xi_1 = \frac{\lambda\eta_1 - \mu\eta_2\tau}{\eta_1^2 - \eta_2^2\tau}, \qquad \xi_2 = \frac{(\mu\eta_1 - \lambda\eta_2)}{\eta_1^2 - \eta_2^2\tau}(-i), \tag{75}$$

where η_1 and η_2 not both zero range independently over all quantities of F so that $\eta_1^2-\eta_2^2\tau\neq 0$. We have therefore

THEOREM 9. *Let i be in F, $i^2=-1$, and λ, μ, ϵ range over all solutions of*

$$\lambda^2 - \mu^2\tau = -\epsilon \tag{76}$$

in F such that $1+\epsilon^2=\tau$ is not the square of any quantity of F. Then every cyclic field of degree eight over F is given by (70), (68), (65), (59), (54), (56), (57), (58) *for every η_1, η_2 not both zero and in F.*

We therefore have only to solve (76). Suppose first that $\mu=0$. Then $\epsilon=-\lambda^2$ and we have proved

THEOREM 10. *Let λ range over all quantities of F such that $1+\lambda^4$ is not the square of any quantity of F. Then* (76) *is satisfied by $\mu=0$, $\epsilon=-\lambda^2$, and defines corresponding cyclic fields.*

Next let $\mu\neq 0$. Define

$$\mu^{-1} = 2\sigma,\ \lambda\mu^{-1} = \rho, \tag{77}$$

so that

$$-\epsilon\mu^{-2} = -4\sigma^2\epsilon = \rho^2 - \tau = \rho^2 - (1+\epsilon^2), \tag{78}$$

$$(\epsilon - 2\sigma^2)^2 - \rho^2 = 4\sigma^4 - 1, \tag{79}$$

and

$$(\epsilon - 2\sigma^2 - \rho)(\epsilon - 2\sigma^2 + \rho) = 4\sigma^4 - 1.$$

Here again we must separate our work into two special cases.

Suppose first that $\epsilon-2\sigma^2-\rho=0$. Then $4\sigma^4=1$, $(2\sigma^2)^2=1$, so that $2\sigma^2=\pm 1$. Moreover if $2\sigma^2=1$ then $(2\sigma)^2=2$, $2\sigma=\mu^{-1}=\pm 2^{1/2}$ so that, since $\lambda=\rho\mu$, we have $\rho=\epsilon-2\sigma^2=\epsilon-1$,

$$\mu = \pm\frac{2^{1/2}}{2},\quad \lambda = (\epsilon - 1)\left(\pm\frac{2^{1/2}}{2}\right), \tag{80}$$

and ϵ ranges over all quantities of F such that $1+\epsilon^2$ is not the square of any quantity of F. Moreover if $2\sigma^2=-1$ then $\mu^{-1}=\pm 2^{1/2}i$, $\rho=\epsilon-2\sigma^2=\epsilon+1$, $\lambda=\rho\mu$,

$$\mu = \pm\frac{2^{1/2}i}{2},\quad \lambda = (\epsilon + 1)\left(\pm\frac{2^{1/2}i}{2}\right). \tag{81}$$

We have therefore proved

THEOREM 11. *Let ϵ range over all quantities of F such that $1+\epsilon^2$ is not the square of any quantity of F. Then if i is in F, $i^2=-1$, and λ, μ are given by either* (80) *or* (81), *so that $2^{1/2}$ is in F, the condition $\lambda^2-\mu^2\tau=-\epsilon$ is satisfied, and Theorem* 9 *defines a set of corresponding cyclic fields of degree eight over F.*

Suppose finally that $\epsilon-2\sigma^2-\rho=\pi\neq 0$. Then $\epsilon-2\sigma^2+\rho=(4\sigma^4-1)\pi^{-1}$ and $2(\epsilon-2\sigma^2)=\pi+(4\sigma^4-1)\pi^{-1}$ while $2\rho=(4\sigma^4-1)\pi^{-1}-\pi$. Also $\lambda=\rho\mu$,

$$\epsilon = \frac{(\pi + 2\sigma^2)^2 - 1}{2\pi},\quad \lambda = \frac{4\sigma^4 - \pi^2 - 1}{4\sigma\pi},\quad \mu = (2\sigma)^{-1}, \tag{82}$$

and we have proved

THEOREM 12. *Let F contain a quantity i such that $i^2=-1$. Then every cyclic field of degree eight over F is given by Theorem* 9 *with λ, μ, ϵ determined by either Theorem* 10 *or* 11 *or by* (82) *as $\pi\neq 0$ in F, $\sigma\neq 0$ in F range over all quantities of F such that $\tau=1+\epsilon^2$ is not the square of any quantity of F.*

7. The case $\tau=-t^2$, t in F. Let $\tau=-t^2$ where t is in F. Then F contains no quantity i such that $i^2=-1$ since otherwise $\tau=(it)^2$ contrary to the fundamental assumption of our work, namely that $F(u)$, $u^2=\tau$, shall be a quadratic field over F. We wish to solve

$$(83)\qquad -\epsilon=[\xi_1^2-(\xi_2 t)^2][\eta_1^2-\eta_2^2\tau],$$

that is, since $\eta_1^2-\eta_2^2\tau\neq 0$,

$$(84)\qquad \xi_1^2-(\xi_2 t)^2=\frac{-\epsilon}{\eta_1^2-\eta_2^2\tau}=R\neq 0,\ -1=\epsilon^2+t^2.$$

Since $\epsilon\neq 0$ we evidently have $\xi_1-\xi_2 t=\pi\neq 0$. Then $\xi_1+\xi_2 t=\pi R^{-1}$ so that

$$(85)\qquad \xi_1=\frac{\pi^2+R}{2\pi},\quad \xi_2=\frac{R-\pi^2}{2t\pi},\quad R=\frac{-\epsilon}{\eta_1^2-\eta_2^2\tau},$$

and we have proved

THEOREM 13. *Let ϵ and t range over all quantities of F such that $-1=\epsilon^2+t^2$ and $1+\epsilon^2=\tau$ is not the square of any quantity of F. Then i is not in F, $i^2=-1$, and every cyclic field of degree eight over F is given by* (68), (65), (59), (54), (56), (57), (58), (85) *when η_1 and η_2 not both zero, $\pi\neq 0$ range independently over all quantities of F.*

8. The case $\tau\neq -t^2$, i not in F. Let -1 be not the square of any quantity of F and let $K=F(i)$, $i^2=-1$, so that $F(i)$ is a quadratic field over K. Our *only remaining case* is the case $-\tau\neq t^2$ for any t of F. This is sufficient to secure the fact that $K(u)$, $u^2=\tau$, is a quadratic field over K,* that is, $F(i, u)$ is a quartic field over F.

For otherwise let $\tau=z^2$, $z=z_1+z_2 i$ where z_1 and z_2 are in F. Then $\tau=z_1^2-z_2^2+2z_1z_2 i$ so that $z_1z_2=0$. But $\tau\neq z_1^2$ in F, by hypothesis. Hence $z\neq z_1$, $z_2\neq 0$ and $z_1=0$. Then $\tau=(z_2 i)^2=-z_2^2$ contrary to hypothesis. We have therefore proved that τ is not the square of any quantity of K, $K(u)$ is a quadratic field over K. We shall now prove

LEMMA. *Let λ and μ be in $K=F(i)$ so that we may write $\lambda=\lambda_1+\lambda_2 i$, $\mu=\mu_1+\mu_2 i$ with λ_1, λ_2, μ_1, μ_2 in F. Let*

$$(86)\qquad \lambda^2-\mu^2\tau=-\epsilon,$$

* This is of course not the field K of preceding sections.

where $\epsilon \neq 0$ *is in* F, $\tau = 1+\epsilon^2$. *Then*

$$\lambda_1\lambda_2 = \tau\mu_1\mu_2, \tag{87}$$

and there exist quantities η_1, η_2 *in* F *and not both zero such that*

$$\lambda_1\eta_2 = \mu_1\eta_1,\ \lambda_2\eta_1 = \mu_2\tau\eta_2,\ \eta_1^2 - \eta_2^2\tau \neq 0. \tag{88}$$

For $-\epsilon = \lambda^2 - \mu^2\tau = [\lambda_1^2 - \lambda_2^2 + \tau(\mu_2^2 - \mu_1^2)] + 2(\lambda_1\lambda_2 - \mu_1\mu_2\tau)i$. Since $-\epsilon$ is in F and i is not in F we have $\lambda_1\lambda_2 - \mu_1\mu_2\tau = 0$ as desired. If $\lambda_1 \neq 0$, $(88)_1$ is satisfied by $\eta_2 = (\lambda_1^{-1}\mu_1)\eta_1$ for every $\eta_1 \neq 0$ of F and

$$\lambda_2\eta_1 - \eta_2\mu_2\tau = [\lambda_2 - (\lambda_1^{-1}\mu_1\mu_2\tau)]\eta_1 = \lambda_1^{-1}\eta_1[\lambda_1\lambda_2 - \mu_1\mu_2\tau] = 0$$

so that (88) is completely satisfied. If $\lambda_1 = 0$, $\mu_2 \neq 0$, then $\mu_1\mu_2\tau = \lambda_1\lambda_2 = 0$ so that $\mu_1 = 0$ and $(88)_1$ is satisfied. Then (88) is satisfied for every $\eta_1 \neq 0$ in F when we take $\eta_2 = (\mu_2\tau)^{-1}\eta_1\lambda_2$. Hence finally let $\lambda_1 = \mu_2 = 0$. Then (88) is merely $\mu_1\eta_1 = \lambda_2\eta_1 = 0$ which is satisfied for any $\eta_2 \neq 0$ in F and by $\eta_1 = 0$. Also $\epsilon \neq 0$ so that, by (86), $\lambda = \lambda_2 i$ and $\mu = \mu_1$ are not both zero, so that necessarily $\eta_1 = 0$.

Consider now the problem of determining a general solution of (71). Suppose we have a solution and then put

$$k_1 = \xi_1 + (\xi_2 i)u,\ k_2 = \eta_1 + \eta_2 u,\ k_3 = \lambda + \mu u = k_1k_2. \tag{89}$$

Equation (89) implies $k_3k_3' = \lambda^2 - \mu^2\tau = -\epsilon = k_1k_1'k_2k_2' = (\xi_1^2 + \xi_2^2\tau)(\eta_1^2 - \eta_2^2\tau)$ and (86) is satisfied where

$$\lambda = \xi_1\eta_1 + \xi_2\eta_2\tau i,\ \mu = \xi_1\eta_2 + \xi_2\eta_1 i. \tag{90}$$

But ϵ is in F and, by the above lemma, $\lambda_1\lambda_2 = \tau\mu_1\mu_2$. Also $\lambda_1 = \xi_1\eta_1$, $\lambda_2 = \xi_2\eta_2\tau$, so that $\lambda_1\eta_2 - \mu_1\eta_1 = \xi_1(\eta_1\eta_2 - \eta_2\eta_1) = 0$, $\lambda_2\eta_1 - \lambda_4\tau\eta_2 = \xi_2\tau(\eta_1\eta_2 - \eta_2\eta_1) = 0$, and (88) is satisfied. Hence every solution of (71) defines a solution of (86) in K for which (87) and (88) are satisfied.

Conversely let (86) be satisfied. By the above lemma, (87), (88) are satisfied. Let η_1, η_2 range over all solutions in F of (88), not both zero, and define k_1, k_2, k_3 by (89) so that if

$$k_1 = \frac{k_3}{k_2} = \frac{\lambda_1 + \mu_1 u}{\eta_1 + \eta_2 u} + \frac{\lambda_2 + \mu_2 u}{\eta_1 + \eta_2 u} i,$$

then

$$\xi_1 = \frac{(\lambda_1\eta_1 - \mu_1\eta_2\tau) + (\mu_1\eta_1 - \lambda_1\eta_2)u}{\eta_1^2 - \eta_2^2\tau} = \frac{\lambda_1\eta_1 - \mu_1\eta_2\tau}{\eta_1^2 - \eta_2^2\tau}$$

is in F by (88). Also

$$\xi_2 u = \frac{\lambda_2 + \mu_2 u}{\eta_1 + \eta_2 u} = \frac{(\lambda_2\eta_1 - \mu_2\eta_2\tau) + (\mu_2\eta_1 - \lambda_2\eta_2)u}{\eta_1^2 - \eta_2^2\tau} = \frac{\mu_2\eta_1 - \lambda_2\eta_2}{\eta_1^2 - \eta_2^2\tau}u,$$

and ξ_2 is in F by (88). Hence (86) determines a set of solutions of (71) and we have proved

THEOREM 14. *Let F contain no quantity i such that $i^2=-1$ and let $\epsilon\neq 0$, λ, μ range over all quantities of $K=F(i)$ such that $\lambda^2-\mu^2\tau=-\epsilon$, ϵ is in F, and $\tau=1+\epsilon^2$, $\pm\tau$ is not the square of any quantity of F. Then if we determine all quantities η_1, η_2 satisfying* (88) *and define $C=F(x)$ by* (55)-(59), (65), (68) *we obtain all cyclic fields C of degree eight over F.*

We therefore need only solve (86). This has already been accomplished in §6. Hence we have, without further proof,

THEOREM 15. *Let t range over all quantities of F such that $\pm(1+t^4)$ is not the square of any quantity of F. Then if $\epsilon=-\lambda^2$, $\lambda=t$ or ti, $\mu=0$, we obtain a solution of $\lambda^2-\mu^2\tau=-\epsilon$ and hence a set of cyclic fields of degree eight over F by the use of Theorem* 14.

Next utilize the *proof* of Theorem 11. If $2\mu=\pm 2^{1/2}$, then either μ is in F and $2^{1/2}$ is in F or $2\mu=\pm ti$, $-t=2^{1/2}i$ is in F, $(-2)^{1/2}$ is in F. Similarly if $2\mu=\pm 2^{1/2}i$ then again $2\mu=\pm t$, ti and either $2^{1/2}$ is in F or $(-2)^{1/2}$ is in F.

THEOREM 16. *Let ϵ range over all quantities of F such that $\pm(1+\epsilon^2)=\pm\tau$ is not the square of any quantity of F and let either $2^{1/2}$ or $(-2)^{1/2}$ be in F but $i=(-1)^{1/2}$ be not in F. Then if either* (80) *or* (81) *is satisfied and λ and μ so defined in $K=F(i)$ we obtain a set of cyclic fields of degree eight over F by the use of Theorem* 14.

We finally use Theorem 12 to state immediately

THEOREM 17. *Let F contain no quantity i, $i^2=-1$. Then every cyclic field of degree eight over F is a cyclic field of Theorems* 13, 15, *or* 16 *or is given by Theorems* 9, 14, *with* (82) *satisfied as $\pi\neq 0$, $\sigma\neq 0$ range over all quantities of $F(i)$ such that ϵ is in F and $\pm(1+\epsilon)^2$ is not the square of any quantity of F.*

UNIVERSITY OF CHICAGO,
CHICAGO, ILL.

ADDITION TO THE NOTE† ON SOME FUNCTIONALS‡

BY

STANISLAW SAKS

7. We need to recall a few known definitions.

Given an abstract space E (i.e., an arbitrary set of elements), a family $\mathfrak{E}$ of sets in E is said to be additive if it satisfies the following conditions:

(i) The empty set (0) belongs to $\mathfrak{E}$.

(ii) If a set X belongs to $\mathfrak{E}$, its complement CX (with respect to the space E) also belongs to $\mathfrak{E}$.

(iii) If $\{X_n\}$ is a sequence of sets belonging to $\mathfrak{E}$, the set $X = \sum X_n$ also belongs to $\mathfrak{E}$.

If $F(X)$ is a finite real-valued function of sets, defined for all sets of an additive family $\mathfrak{E}$, and if

$$F\left(\sum_n X_n\right) = \sum_n F(X_n) \tag{7.1}$$

for any finite sequence $\{X_n\}$ of sets of $\mathfrak{E}$, of which no two have points in common, then $F(X)$ is called an additive function of sets of $\mathfrak{E}$. If (7.1) holds for any finite or infinite sequence $\{X_n\}$ of sets belonging to $\mathfrak{E}$, of which no two have points in common, then $F(X)$ is said to be a completely additive function of sets of $\mathfrak{E}$.

In this paragraph we assume that $\mathfrak{R}^*$ is an additive family in the space E, and $\mu(X) \geqq 0$ is a completely additive and finite-valued function of sets of $\mathfrak{R}^*$. The sets X belonging to $\mathfrak{R}^*$ are called measurable, $\mu(X)$ being the measure of X. A measurable set X is a singular set if for any measurable subset Y of X either $\mu(Y)=0$ or $\mu(X-Y)=0$.

An additive function $F(X)$ of measurable sets is absolutely continuous if $F(X)=0$ whenever X is of measure zero. This together with the property of being completely additive, is equivalent to the statement that for any $\epsilon>0$ there exists an $\eta>0$ such that $\mu(X)<\eta$ implies $|F(X)|<\epsilon$.

The family $\mathfrak{R}^*$ of measurable sets may be regarded as a metric complete space with the distance defined by §

† This volume, pp. 549–556. In the present addition we extend the results of §2 to completely additive functions of sets in an abstract space. The author is indebted to Professor Tamarkin for criticisms.

‡ Presented to the Society, April 14, 1933; received by the editors February 16, 1933.

§ This definition corresponds to that of distance in the space R of characteristic functions of §2.

$$d(X_1, X_2) = \mu(X_1 - X_1X_2) + \mu(X_2 - X_1X_2). \tag{7.2}$$

If two measurable sets differ by subsets of measure zero they are regarded as the same elements of the space $\mathfrak{R}^*$. Any completely additive and absolutely continuous function of measurable sets may be regarded as a continuous functional on the metric space $\mathfrak{R}^*$.

LEMMA 1. *If A is a measurable set of positive measure, then, for any positive number ϵ, the set A contains either a singular set of measure $>\epsilon$ or a measurable set of positive measure $\leqq \epsilon$.*

Suppose that A contains neither a singular set of measure $>\epsilon$, nor a measurable set of positive measure $\leqq\epsilon$. Then there will exist a measurable subset A_1 of A such that $0<\mu(A_1)<\mu(A)$. The set $A-A_1$ must be a non-singular set of measure $>\epsilon$, and, by the same argument, $A-A_1$ contains a measurable subset A_2 such that $0<\mu(A_2)<\mu(A-A_1)$. By repeating this process we obtain an infinite sequence of measurable sets $\{A_n\}$ of positive measure, of which no two have points in common. Since the series

$$\sum_{n=1}^{\infty}\mu(A_n) = \mu\left(\sum_{n=1}^{\infty}A_n\right)$$

converges, for n sufficiently large, we have $0<\mu(A_n)<\epsilon$. This, however, contradicts the assumption that A contains no measurable set of measure $\leqq\epsilon$.

LEMMA 2. *Given an arbitrary number $\epsilon>0$, the space E may be expressed as the sum of a finite number of measurable sets $E_1, E_2, \cdots, E_p$ such that $E_iE_j=0$ for $i\neq j$, while each E_i is either a singular set or a set of measure $\leqq\epsilon$.*

We observe that for an arbitrary pair of singular sets, either one of them contains the other, with the possible exception of a set of measure zero, or else their common part is of measure zero. Since $\mu(E)<\infty$, on the basis of this remark we can find a finite sequence of singular sets $E_1, E_2, \cdots, E_n$ of measure $>\epsilon$ such that

$$E_iE_j = 0 \text{ for } i \neq j, \tag{7.3}$$

while the set

$$A = E - \sum_{i=1}^{m} E_i \tag{7.4}$$

contains no singular set of measure $>\epsilon$.

Let X be any measurable set and let $\lambda(X)$ denote the least upper bound of the measures of all measurable subsets Y of X such that $\mu(Y)\leqq\epsilon$. It follows from Lemma 1 that $0<\lambda(X)\leqq\epsilon$ for any measurable set $X\subset A$ of positive measure. Hence, by induction, we can determine a sequence $\{X_i\}$ of measurable subsets of A such that

$$X_iX_j = 0 \text{ for } i \neq j, \tag{7.5}$$

$$\epsilon \geqq \mu(X_{n+1}) \geqq \tfrac{1}{2}\lambda\left(A - \sum_{i=1}^{n} X_i\right) \qquad (n = 1, 2, \cdots). \tag{7.6}$$

Upon putting

$$X_0 = A - \sum_{i=1}^{\infty} X_i$$

from (7.6) we have

$$\lambda(X_0) \leqq \lambda\left(A - \sum_{i=1}^{n} X_i\right) \leqq 2\mu(X_{n+1}) \qquad (n = 1, 2, \cdots). \tag{7.7}$$

Since, by (7.5),

$$\sum_{i=1}^{\infty} \mu(X_i) \leqq \mu(A) < \infty\ , \tag{7.8}$$

the series (7.8) converges and $\lim_n \mu(X_n) = 0$. Thus we infer from (7.7) that $\lambda(X_0) = 0$, whence also $\mu(X_0) = 0$. Let now h be a positive integer such that

$$\mu\left(\sum_{n=h+1}^{\infty} X_n\right) = \sum_{n=h+1}^{\infty} \mu(X_n) \leqq \epsilon, \tag{7.9}$$

and let

$$E_{m+1} = X_1, \ \cdots, E_{m+h} = X_h, E_{m+h+1} = X_0 + \sum_{n=h+1}^{\infty} X_n.$$

These sets, by (7.6) and (7.9), are of measure $\leqq \epsilon$, and by (7.5) no two of them have points in common. Hence the sequence $E_1, E_2, \cdots, E_{m+h+1}$ satisfies the conditions of Lemma 2.

8. We now are able to generalize Theorems 1 and 2 of §2.

THEOREM 5. *Let $\{F_n(X)\}$ be a sequence of completely additive and absolutely continuous functions of measurable sets. If this sequence converges for any set belonging to a class of the second category in the space $\mathfrak{R}^*$, then the functions $F_n(X)$ are equally absolutely continuous*† *and the sequence $\{F_n(X)\}$ converges for any measurable set $X \subset E - (E_1 + E_2 + \cdots + E_m)$ where $\{E_i\}$ is a finite sequence of singular sets.*

Consequently, if $\{F_n(X)\}$ converges for any measurable set X, the limit function is again a completely additive and absolutely continuous function of measurable sets in E.

† That is, to every $\epsilon > 0$ there corresponds an $\eta > 0$ which depends only on ϵ, such that $|F_n(X)| \leqq \epsilon$ for $n = 1, 2, \cdots$ and for any set X of measure $\leqq \eta$.

The fact that the functions $F_n(X)$ are equally absolutely continuous can be established in exactly the same fashion as in Theorem 1, §2, if we interpret the functions $F_n(X)$ as continuous functionals in the metric complete space $\mathfrak{R}^*$. Now, since by assumption the sequence $\{F_n(X)\}$ converges for any X belonging to a set of the second category in $\mathfrak{R}^*$, there exists in $\mathfrak{R}^*$ a sphere, say $\mathfrak{K}(A_0; r)$, such that $\{F_n(X)\}$ converges for each X of a set everywhere dense in $\mathfrak{K}(A_0; r)$. But the functionals $F_n(X)$ are equally continuous in $\mathfrak{R}^*$, hence the sequence $\{F_n(X)\}$ converges everywhere in the sphere $\mathfrak{K}(A_0; r)$.

Now let

$$E = \sum_{i=1}^{p} E_i$$

be a representation of the space E mentioned in Lemma 2. We may assume that the sets $E_1, E_2, \cdots, E_m$ are singular while the sets $E_{m+1}, \cdots, E_p$ are of measure $\leqq r$.

Let X be an arbitrary measurable set contained in $\sum_{m+1}^{p} E_i$. Then

$$X = \sum_{i=m+1}^{p} XE_i. \tag{8.1}$$

Each set XE_i, $i=m+1, \cdots, p$, is of measure $\leqq r$. Consequently the sets A_0+XE_i and $A_0-A_0XE_i$, $i=m+1, \cdots, p$, are elements of the sphere $\mathfrak{K}(A_0;r)$ and both sequences $\{F_n(A_0+XE_i)\}$, $\{F_n(A_0-A_0XE_i)\}$ converge. Thus the sequence

$$F_n(XE_i) = F_n(A_0 + XE_i) - F_n(A_0 - A_0XE_i)$$

also converges for $i=m+1, \cdots, p$. Hence, by (8.1), the sequence $\{F_n(X)\}$ converges for any measurable set X contained in $E-(E_1+\cdots+E_m)$ where $E_1, \cdots, E_m$ are singular sets.

THEOREM 6. *If $\{F_n(X)\}$ is a sequence of completely additive and absolutely continuous functions of measurable sets and if*

$$\overline{\lim_n}\ |F_n(X)| < \infty \tag{8.2}$$

for any set X belonging to a class of the second category in the space $\mathfrak{R}^*$, then there exists a fixed constant M such that

$$|F_n(X)| < M \tag{8.3}$$

for any measurable set $X \subset E-(E_1+\cdots+E_m)$ where $\{E_i\}$ is a finite sequence of singular sets in E.

Consequently, if the inequality (8.2) *holds for every measurable set X, there exists a constant M such that* (8.3) *holds for all measurable sets X in E.*

Let $\mathfrak{N}_k^*$ be the aggregate of sets X such that

$$|F_n(X)| \leqq k \qquad (n = 1, 2, \cdots).$$

By assumption the class $\sum_1^\infty \mathfrak{N}_k^*$ is of the second category in the space $\mathfrak{N}^*$. By the continuity of the functionals $F_n(X)$, the sets $\mathfrak{N}_k^*$ are closed (in the space $\mathfrak{N}^*$). Hence, for some value $k=k_0$, $\mathfrak{N}_{k_0}^*$ contains a sphere, say $\mathfrak{K}(A_0;r)$.

We now introduce the same representation of the space

$$E = \sum_{i=1}^{p} E_i$$

as in the proof of Theorem 5. Let X be an arbitrary measurable set. Since, for $i=m+1, \cdots, p$, the sets XE_i are of measure $\leqq r$, the sets A_0+XE_i and $A_0-A_0XE_i$ belong to the sphere $\mathfrak{K}(A_0;r)$. Thus

$$|F_n(A_0 + XE_i)| \leqq k_0, \ |F_n(A_0 - A_0XE_i)| \leqq k_0,$$

and

$$|F_n(XE_i)| = |F_n(A_0 + XE_i) - F_n(A_0 - A_0XE_i)| \leqq 2k_0.$$

Hence, for any measurable set $X \subset E-(E_1+ \cdots +E_m)$ we have

$$|F_n(X)| = \left|F_n\left(\sum_{i=m+1}^{p} XE_i\right)\right| \leqq 2(p-m)k_0,$$

which completes the proof of Theorem 6.

9. Theorems 5 and 6 contain the corresponding two theorems which have been stated recently by Nikodym.†

I. If $\mathfrak{E}$ is an additive family of sets in an abstract space E, and if the sequence $\{F_n(X)\}$ of completely additive functions of sets of $\mathfrak{E}$ converges for every set X of $\mathfrak{E}$, then the limit function is also a completely additive function of sets of $\mathfrak{E}$.

II. If $\mathfrak{E}$ is an additive family of sets in E and if the sequence $\{F_n(X)\}$ of completely additive functions of sets of $\mathfrak{E}$ is bounded for every set X of $\mathfrak{E}$, then there exists a constant M such that $|F_n(X)| \leqq M$ for $n=1, 2, \cdots$ and for all $X \subset \mathfrak{E}$.

In order to reduce these theorems to Theorems 5 and 6 respectively we merely have to introduce a measure $\mu(X)$ for the family $\mathfrak{E}$, with respect to which the functions $F_n(X)$ would be absolutely continuous. This can be achieved by putting, for each set $X \subset \mathfrak{E}$,

† O. Nikodym, *Sur les suites des fonctions parfaitement additives d'ensembles abstraits*, Comptes Rendus, vol. 192 (1931), pp. 727–728. The proofs of the results stated in that note will appear in the Monatshefte für Mathematik und Physik.

$$\mu(X) = \sum_{n=1}^{\infty} \frac{V_n(X)}{2^n[V_n(E) + 1]}, \tag{9.1}$$

where $V_n(X)$ denotes the absolute variation of $F_n(X)$ on the set X. Since each $V_n(X)$ is a non-negative and completely additive function† of sets of $\mathfrak{E}$ the series (9.1) converges and $\mu(X) \geqq 0$ is a completely additive and finite-valued function of sets of $\mathfrak{E}$. Hence $\mu(X)$ may be taken as a measure in E and, since $F_n(X)=0$, $n=1, 2, \cdots$, for every set X of $\mathfrak{E}$ such that $\mu(X)=0$, the functions $F_n(X)$ are absolutely continuous with respect to this measure. Thus the theorems of Nikodym are reduced to our Theorems 5 and 6.

† See for instance H. Hahn, *Theorie der reellen Funktionen*, 1921, Chapter VI. The absolute variation is called there (p. 400) "absolute Summe."

University of Warsaw,
Warsaw, Poland

A SECOND CORRECTION

BY

EDWARD V. HUNTINGTON

In my paper in the present volume these Transactions (vol. 35, pp. 274–304 and 557–558), the Example 5 on page 304 is erroneous and Postulate 5 on page 301 is redundant. That is, Postulates 1, 2, 3, 4, 6, 7 (without 5) form a set of independent postulates for the "informal" system of *Principia Mathematica*.

The proof of 5 from 1, 2, 3, 4, 6, 7 is as follows.*

6a. *If $a+b$ is in T and a not in T, then b is in T.* (From 6.)

6b. *If a not in T and b not in T, then $a+b$ not in T.* (From 6.)

7a. *If a is in T, then a' is not in T.* (From 7.)

3a. *If b is in T, then $a+b$ is in T.*

For, by 7a, b' is not in T. But by 3, $b'+(a+b)$ is in T. Hence by 6a, $a+b$ is in T.

4a. *If b is in T, then $b+a$ is in T.*

For, by 3a, $a+b$ is in T, whence by 7a, $(a+b)'$ is not in T. But by 4, $(a+b)'+(b+a)$ is in T. Hence by 6a, $b+a$ is in T.

5a. *If a is not in T, then a' is in T.*

For, suppose a' not in T. Then by 6b, $a'+a$ not in T, whence by 6b, $a'+(a'+a)$ not in T, contrary to 3.

5. *If a, b, etc. are in K, then $(b'+c)'+[(a+b)'+(a+c)]$ is in T.*

Case 1; *a in T.* By 4a, $a+c$ is in T. Hence the theorem, by 3a (twice).

Case 2: *b in T.* By 7a, b' is not in T. If c is in T, then by 3a, $a+c$ is in T, whence the theorem, by 3a (twice). If c is not in T, then by 6b, $b'+c$ is not in T, whence by 5a, $(b'+c)'$ is in T, whence the theorem, by 4a.

Case 3: *a not in T and b not in T.* By 6b, $a+b$ not in T, whence by 5a, $(a+b)'$ is in T. Hence the theorem, by 4a and 3a.

The proof is thus complete. It can also be shown that 1, 2, 3a, 4a, 5a, 6a, 7 form a set of independent postulates equivalent to the set 1, 2, 3, 4, 6, 7.

---

* For valuable suggestions in this connection I am indebted to Professor Alonzo Church and Dr. K. E. Rosinger.

HARVARD UNIVERSITY,
CAMBRIDGE, MASS.
May 29, 1933.

CORRECTION TO A PAPER ON THE MOORE-KLINE PROBLEM*

BY

LEO ZIPPIN

It has been brought to my attention by Mr. N. E. Steenrod that the lemma of page 708 in the paper referred to is in error. The final assertion of the proof is false. It is therefore necessary to point out that the paper is not "disturbed" by this fault. For if one requires that the P_n, $n=1, 2, \cdots$, of the lemma be *arcs* then the (altered) lemma does hold, since it is true that the connected sum of a perfect continuous curve and an arc is "perfect." One verifies that this restricted lemma is sufficient for the uses of the paper.

---

* Published under that title, these Transactions, vol. 34, pp. 705–721.

PRINCETON, UNIVERSITY,
PRINCETON, N. J.